PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/82, 15/12, C07K 14/435, G01N 33/52, 33/53

A1

(11) International Publication Number: WO 96/27675

(43) International Publication Date: 12 September 1996 (12.09.96)

(21) International Application Number: PCT/GB96/00481

(22) International Filing Date: 4 March 1996 (04.03.96)

(30) Priority Data:
9504446.7 6 March 1995 (06.03.95) GB

(71) Applicant *(for all designated States except US)*: MEDICAL RESEARCH COUNCIL [GB/GB]; 20 Park Crescent, London W1N 4AL (GB).

(72) Inventors; and

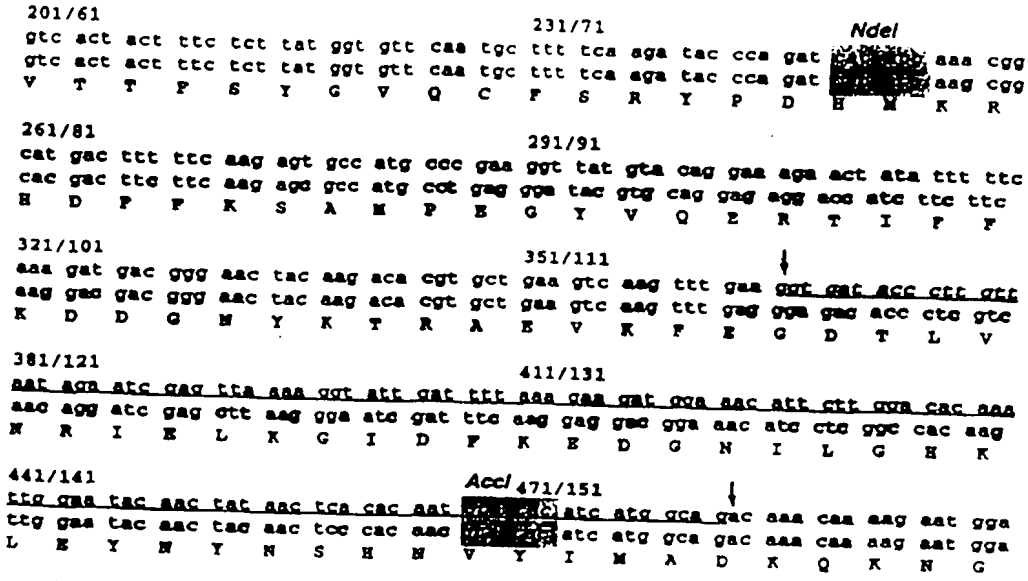
(75) Inventors/Applicants *(for US only)*: HASELOFF, James, Phillip [AU/AU]; Glebe Cottage, Church Lane, Comberton, Cambridgeshire CB3 7ED (GB). HODGE, Sarah [GB/GB]; 15 St Luke's Street, Cambridge CB4 3DA (GB). PRASHER, Douglas [US/US]; 72 Goeletta Drive, East Falmouth, MA 02536-3964 (US). SIEMERING, Kirby [AU/AU]; 8 Portugal Place, Cambridge CB5 8AF (GB).

(74) Agent: KEITH W. NASH & CO.; 90-92 Regent Street, Cambridge CB2 1DP (GB).

(81) Designated States: AU, CA, JP, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.

(54) Title: JELLYFISH GREEN FLUORESCENT PROTEIN (GFP) EXPRESSION IN PLANTS



top sequence = Aequoria victoria GFP, lower sequence = mGFP4
Intron sequences are underlined and the cryptic splice junctions are arrowed.
Mutated nucleotides are shown outlined. Nucleotide and amino acid numbering starts at the initiation codon.

(57) Abstract

Disclosed is a DNA sequence encoding Green Fluorescent Protein (GFP), the sequence being modified relative to the wild type sequence so as to allow for more efficient expression in a plant cell of a functional GFP polypeptide. Also disclosed is a modified GFP polypeptide comprising an amino acid substitution, relative to the wild type protein sequence, which exhibits useful characteristics when expressed in many different types of host cell.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgystan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic			SE	Sweden
CG	Congo	KR	Republic of Korea	SG	Singapore
CH	Switzerland	KZ	Kazakhstan	SI	Slovenia
CI	Côte d'Ivoire	LI	Liechtenstein	SK	Slovakia
CM	Cameroon	LK	Sri Lanka	SN	Senegal
CN	China	LR	Liberia	SZ	Swaziland
CS	Czechoslovakia	LT	Lithuania	TD	Chad
CZ	Czech Republic	LU	Luxembourg	TG	Togo
DE	Germany	LV	Latvia	TJ	Tajikistan
DK	Denmark	MC	Monaco	TT	Trinidad and Tobago
EE	Estonia	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	UG	Uganda
FI	Finland	ML	Mali	US	United States of America
FR	France	MN	Mongolia	UZ	Uzbekistan
GA	Gabon	MR	Mauritania	VN	Viet Nam

JELLYFISH GREEN FLUORESCENT PROTEIN (GFP) EXPRESSION IN PLANTS

Field of the Invention

This invention relates to improvements in gene expression, especially improvements in expression of the Green Fluoresecent Protein (GFP) gene, and to a method of detecting the presence and/or expression in a host of a gene of interest.

Background of the Invention

Genes encoding β-glucuronidase and β-galactosidase have been used as reporters for gene expression in plants. Using these reporter genes, transformed tissues or patterns of gene expression can be identified histochemically, but this is generally a destructive test and is not suitable for assaying primary transformants, nor for following the time course of gene expression in living plants, nor as a means of rapidly screening segregating populations of seedlings. There is thus a general need for improved reporters of gene expression, but especially reporter genes for use in plants. Candidates might be found among proteins having intrinsic fluorescence.

Proteins with high intrinsic fluorescence are involved in photosynthesis and bioluminescence, and in most cases possess a protein-bound chromophore. For example, the highly fluorescent phycobiliproteins require complex tetrapyrrole groups, and the blue and yellow fluorescent proteins from *Vibrio fischeri* must bind lumazine and flavin mononucleotide, respectively. This requirement for an external chromophore complicates the use of these proteins as reporters for gene expression. However, the green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* does not share this requirement for an external chromophore.

Aequorea victoria are brightly luminescent, with light appearing as glowing points around the margin of the jellyfish umbrella. Light arises from yellow tissue masses which each consist of about 6000-7000 photogenic cells (Davenport & Nichol, 1955 Proc. Roy. Soc. Ser. B 144, 399-411). The cytoplasm of these cells is densely packed with fine granules

of about 0.2μm diameter which are enclosed by a unit membrane and contain the components necessary for bioluminescence (Anderson & Cormier, 1973 J. Biol. Chem. *248*, 2937-2943). The components include a Ca^{++} activated photoprotein, aequorin, that emits blue-green light, and an accessory green fluorescent protein (GFP) which accepts energy from aequorin and re-emits it as green light.

GFP is an extremely stable protein of 238 amino acids. The fluorescent properties of the protein are unaffected by prolonged treatment with 6M guanidine HCl, 8M urea or 1% SDS, and two day treatment with various proteases such as trypsin, chymotrypsin, papain, subtilisin, thermolysin and pancreatin at concentrations up to 1 mg/ml fail to alter the intensity of GFP fluorescence. GFP is stable in neutral buffers up to 65°C, and displays a broad range of pH stability from 5.5 to 12 (Bokman & Ward, 1981 Biochem. Biophys. Res. Comm. *101*, 1372-1380). The protein is intensely fluorescent, with a quantum efficiency of approximately 80% and molar extinction coefficient of around $4.5x10^4$. GFP absorbs light maximally at 395 nm and has a smaller absorbance peak at 475nm, and fluorescence emission peaks at 509nm, with a shoulder at 540nm (Morise *et al.*, 1974 Biochemistry *13*, 2656-2662). Researchers have successfully cloned and sequenced both the cDNA and genomic DNA sequences coding for *A. victoria* GFP (Prasher *et al.*, 1992 Gene *111*, 229-233).

The fluorescence of GFP has been well characterised (Inouye & Tsuji, 1994 FEBS Letters *13817*, 277-280 and FEBS Letters *14472*, 211-214) and appears to be due to a unique covalently-attached chromophore which is formed post-translationally by cyclisation and oxidation of the residues 65-67 (Ser-Tyr-Gly) within the protein (Cody *et al.*, 1993 Biochemistry *32*, 1212-1218; Heim *et al.*, 1994 PNAS *91*, 12501-12504). Several genomic and cDNA clones of *gfp* have been obtained from a population of *A. victoria*. The *gfp* gene contains at least three introns, and the sequences derived from the cDNA have been used for protein expression studies in *Escherichia coli*, *Caenorhabditis elegans* (Chalfie *et al.*, 1994 Science *263*, 802 *et seq.*) and *Drosophila melanogaster* (Wang & Hazelrigg, 1994 Nature *369*, 400 *et seq.*). Fluorescent protein was produced in these different cell types and there appears to be little requirement for specific additional factors for post-translational modification of the protein, which may be

autocatalytic or require common factors. Recently, modified forms of GFP have also been made and studied, which forms have rather different fluorescence characteristics compared to the wild type (Heim *et al.*, 1994 PNAS *91*, 12501-12504; Delagrave *et al.*, 1995 Bio/Technology *13*, 151 *et seq.*; and Heim *et al.*, 1995 Nature *373*, 663-664).

Although GFP has some advantages as a fluorescent reporter molecule, expression has been reported to be problematic in some experimental systems (Cubitt *et al.*, 1995 Trends Biochem. Sci. *20*, 448-455). Expression of GFP in mammalian cells has been described as highly variable (Rizutto *et al.*, 1995 PNAS *92*, 11899-11903); Kaether & Gerdes 1995, FEBS Lett. *369*, 267-271; Pines 1995, Trends Genet. *11*, 326-327) often requiring a strong promoter and decreased incubation temperature for good results (Ogawa *et al.*, 1995 PNAS *92*, 11899-11903). Other researchers have found that a lower incubation temperature also favours the development of fluorescence during expression of GFP in bacteria (Heim *et al.*, 1994 PNAS *91*, 12501-12504; Webb *et al.*, 1995 J. Bact. *177*, 5906-5911) and yeast (Lim *et al.*, 1995 J. Biochem. *118*, 13-17). In yeast, this phenomenon has been attributed primarily to more efficient maturation of GFP to the fluorescent form at lower temperatures. The present inventors have sought to express modified forms of GFP in various hosts.

Summary of the Invention

In a first aspect the invention provides a DNA sequence encoding Green Fluorescent Protein (GFP), the sequence being modified relative to the wild type sequence so as to allow for more efficient expression in a plant cell of a functional GFP polypeptide. Preferably the modified sequence is capable of efficient expression in a dicotyledonous plant, such as *Arabidopsis*.

The term "GFP" as used herein is intended to refer to a polypeptide possessing many of the properties of the naturally occurring protein, and particularly exhibiting intrinsic fluorescence. As will be apparent to those skilled in the art, the polypeptide need not necessarily fluoresce in the "green" part of the visible spectrum, as the fluorescence properties of the polypeptide (including the wavelength of fluorescence) may be

substantially altered by one or more mutations.

The GFP polypeptide will generally have substantially the amino acid sequence of the wild type protein (as disclosed by Prasher *et al.*), but will preferably comprise one or more amino acid differences defined, and discussed in greater detail, below.

The GFP-coding DNA sequence is advantageously modified so as to reduce the probability of an RNA sequence transcribed therefrom being subject to erroneous splicing in a plant cell. The DNA sequence is conveniently modified so as to comprise a plurality of nucleotide substitutions relative to the wild type sequence, which substitutions serve to reduce, or preferably entirely prevent, excision from the transcribed RNA of the portion corresponding to nucleotides 400-483 of the DNA sequence. It has surprisingly been found by the present inventors that this portion of the sequence tends to be recognised in plant cells (particularly dicotyledenous plants) as an intron, which is therefore excised by splicing of the RNA.

Preferably the nucleotide substitutions in the DNA sequence serve to decrease the A/U% content of the transcribed RNA, which is believed to decrease the likelihood of the sequence being treated by a plant cell as representing an intron. Desirably the substitutions particularly decrease the A/U% content of the region corresponding to nucleotides 400-483. Conveniently the nucleotide substitutions are such as to preserve the amino acid sequence of the encoded polypeptide substantially unchanged in the portion encoded by nucleotides 400-483. Other substitutions may advantageously be made to decrease the similarity between the GFP RNA sequence and the plant intron recognition consensus sequence (see Figure 2).

In addition to nucleotide substitutions in the portion 400-483 of the DNA sequence, a number of other modifications may advantageously be made. For example, substitutions may be made downstream (i.e. 3') and/or upstream (i.e. 5') of nucleotides 400-483, which substitutions will conveniently serve to further reduce the A/U content of the transcribed RNA.

The invention also includes within its scope RNA sequences capable of being transcribed from the modified DNA sequence (i.e. RNA sequences transcribed from the modified DNA sequence, or an RNA having a sequence such that it could be synthesised by transcription from the modified DNA sequence).

Advantageously a number of other modifications, in addition to those specified above, may also be made to the GFP-coding sequence. For example, the sequence will typically further comprise transcription and translation signals (e.g. promoters, enhancers) and/or localisation signals recognised in plants. Localisation signals may direct the expressed polypeptide to: the nucleus (e.g. SV40 large T antigen localisation signal, particularly in combination with other polypeptide sequences, which have been found to increase the efficiency of the signalling); mitochondria (e.g. cytochrome C oxidase subunit IV); endoplasmic reticulum - ER (e.g. the signal sequence from carboxypeptidase Y, or that from *Arabidopsis* basic chitinase); or microbodies (e.g. peroxisomes). It may even be possible to target the modified GFP polypeptide to the plant cell wall. Localisation of GFP may be highly desirable when expression occurs at high levels, so as to minimise possible toxicity to host cells.

The sequence may advantageously be further modified in accordance with the manner described in the prior art (e.g. as disclosed by Heim *et al.*, 1994, 1995, or Delagrave *et al.*, 1995, as cited previously). Whilst in general the DNA sequence is modified in such a way as to preserve the wild type amino acid sequence, it has been found that amino acid changes at specific residues are in fact desirable. In particular, the sequence may be modified so as to comprise an amino acid substitution at one or both of amino acid residues 163 and 175. Changes at these positions are found to alter the characteristics of the polypeptide in an unexpected and favourable manner.

In particular, amino acid substitutions at residue 163 and/or 175 (valine and serine respectively, in the wild type protein) have favourable effects on the characteristics of the GFP polypeptide when expressed in many different host cells (e.g. bacterial, yeast etc.), and such substitutions may advantageously be included independently of any modification of the DNA sequence made for increased efficiency of expression in plants.

In a second aspect therefore the invention provides a modified GFP polypeptide comprising an amino acid substitution relative to the wild type protein at residue 163 and/or 175. Substitution at either residue, in isolation, has surprisingly been found to increase the thermotolerance of the polypeptide. Maximal thermotolerance is obtained by causing substitution at both residues.

Advantageously valine 163 is substituted by alanine, or by a related amino acid (i.e. those having an aliphatic side chain: glycine, leucine and isoleucine; or those having an aliphatic hydronyl side chain: serine and threonine).

Advantageously serine 175 is substituted by glycine, or by a related amino acid (i.e. those having an aliphatic side chain: alanine, leucine and isoleucine; or those having an aliphatic hydroxyl side chain: serine and threonine).

It is preferred that the modified GFP polypeptide comprises substitutions at both residues, conveniently 163→alanine and 175→glycine.

The modified GFP may additionally comprise other sequence differences relative to the wild type protein, particularly in, or immediately adjacent to, residues 65-67 (which residues give rise to the chromophore).

Such a nucleic acid sequence is useful for example, as a marker, or as a reporter gene, in a wide variety of host cells (e.g. mammalian, bacterial, fungal, yeast or plant cells). Conveniently, the nucleic acid sequence may be further modified for expression in a particular host cell. For example, where the thermotolerant GFP-coding sequence is to be expressed in a plant cell it will conveniently be modified in accordance with the first aspect of the invention.

In a third aspect, the invention provides a nucleic acid construct comprising a nucleic acid sequence in accordance with the first aspect of the invention. In particular the construct is preferably an expression vector, comprising one or more regulatory signals (such as promoters etc.) and is preferably suitable for use in a plant cell. The construct will

desirably include one or more restriction endonuclease sites, suitable for the insertion into the construct of other nucleic acid sequences, which in a preferred embodiment may be inserted in frame with the sequence of the invention.

The invention also provides a host cell, conveniently a plant cell, into which has been introduced a sequence in accordance with the first aspect of the invention.

Processes for introducing DNA into plant cells (typically by transformation) are not 100% efficient. Accordingly, it is generally desirable for the DNA introduced into the plant cell to confer one or more distinctive characteristics upon the plant cell, which characteristic(s) serve to mark those cells which have taken up the DNA. The fluorescent properties of GFP constitute such a distinctive characteristic ("marker"). In a preferred embodiment the invention thus provides a plant cell transformation vector comprising the sequence of the invention. Further, the invention provides a method of screening plant cells, comprising introducing into at least some of a plurality of plant cells a DNA construct comprising a sequence in accordance with the invention, maintaining the cells under suitable conditions for an appropriate length of time so as to allow expression of a modified GFP from the construct, and selecting those cells which exhibit GFP-mediated fluorescence. "Suitable conditions" and "an appropriate length of time" are well known to those skilled in the art from standard texts.

In a preferred embodiment, the vector further comprises a sequence of interest which, preferably, is present in frame with the modified GFP-coding sequence.

In a fourth aspect the invention thus provides a method of detecting the expression in a plant of a sequence of interest, comprising causing the sequence of interest to be present in frame with a modified GFP-coding sequence in accordance with the first aspect of the invention so as to form a modified GFP/sequence of interest fusion, introducing the fusion into a plant, and monitoring the fluorescence thereof. GFP-mediated fluorescence is thus an indicator of expression of the sequence of interest.

In yet another aspect, the invention provides a nucleic acid construct comprising a

sequence in accordance with the second aspect of the invention. The nucleic acid construct will desirably have many of the features of the nucleic acid construct in accordance with the third aspect of the invention. It will be apparent however that the construct may be useful in many different types of host cell, and may be constructed accordingly.

The invention will now be further described by way of illustration and with reference to the accompanying drawings, in which:

Figure 1A shows the sequences introduced, via PCR, flanking the GFP-coding sequence;

Figure 1B is a confocal micrograph of transformed yeast cells expressing GFP;

Figure 2A is a photograph showing agarose gel electrophoresis analysis of PCR products;

Figure 2B is a schematic illustration of the portion of DNA not represented in the mis-spliced mRNA produced in plants from the wild type GFP-coding sequence;

Figure 3A is a photograph showing the DNA sequence determination of the reverse transcript produced from mis-spliced mRNA;

Figure 3B is a comparison between a portion of the GFP wild type sequence and a plant intron consensus sequence;

Figure 4 shows a comparison of part of the wild type *A. victoria* GFP sequence with a modified GFP-coding sequence in accordance with the invention;

Figure 5 is a series of confocal micrographs (at different magnification) showing parts of a plant expressing a modified GFP-coding sequence in accordance with the invention;

Figure 6 shows a comparison of part of three modified GFP-coding sequences in accordance with the invention, together with the amino acid sequences encoded thereby;

Figure 7 is a graph of relative fluorescence (arbitary units) against time (minutes) for *E. coli* strains expressing modified GFP (open squares) or modified, mutated GFP (filled circles);

Figure 8 is a photograph of a Western blot, probed with anti-GFP antibody;

Figure 9 is a bar chart showing the amount of fluorescence associated with cultures expressing modified GFP (open columns) on modified, mutated GFP (shaded columns) incubated at four different temperatures;

Figure 10 is a graph of fluorescence against time (minutes) for yeast cultures at 25°C or 37°C, the cultures having been grown initially in anaerobic conditions, with oxygen introduced at time zero;

Figure 11 is a picture of a Western blot showing expression of modified GFP or modified, mutated GFP (GFPA) by *E. coli* cultures at 25 or 37°C, with comparison between soluble and insoluble culture fractions;

Figure 12 is a graph of absorbance against wavelength for soluble (filled circles) or insoluble (open circles) GFP;

Figure 13 is a picture of yeast cultures, grown at 25 or 37°C and expressing modified GFP, or modified, mutated GFP (GFPA);

Figure 14 is a graph of fluorescence against wavelength (nm), showing the excitation spectra (squares) and emission spectra (circles) respectively, of modified GFP (solid lines) and two mutated forms of modified GFP, GFPA (dashed lines) and GFP5 (dotted lines);

Figure 15 is a comparison of the nucleotide sequence of wild-type *gfp* and a modified gene *m-gfp5*, and the polypeptides encoded thereby. Nucleotide sequence differences are shown in bold. The m-GFP5 amino acid sequence is shown beneath the nucleotide

sequence. The three amino acid differences between the encoded polypeptides are indicated:

Figure 16 shows the sequence of another modified *m-gfp* gene, termed *m-gfp5-ER*, and the amino acid sequence of the polypeptide encoded thereby; and

Figure 17 shows a number of confocal micrographs (A-H) of *Arabidopsis* seedlings expressing modified *gfp* genes in accordance with the invention.

Examples

Example 1

Construction of a *gfp* expression cassette

A synthetic *gfp* gene was constructed using the polymerase chain reaction (PCR). The plasmid pGFP10.1 (described by Prasher *et al.*, 1992 Gene *111*, cited above) contains a cloned *A. victoria gfp* cDNA, and was used as template for PCR amplification (with *Thermococcus litoralis* Vent polymerase) with synthetic oligomer primers which were used to incorporate new sequences flanking the GFP coding sequence.

The sequence of the primer oligonucleotides was:
GGCGGATCCAAGGAGATATAACAATGAGTAAAGGAGAAGAACTTTTCACT (Seq. ID No. 1) and GGCGAGCTCTTATTTGTATAGTTCATCCATGCC (Seq. ID No. 2).

The newly-incorporated sequences are shown in Figure 1A. Referring to Figure 1A, the sequence existing in pGFP10.1 is shown italicised. The added sequences are shown in normal type. These included: recognition sites for the restriction endonucleases *Bam*HI and *Sac*I placed at the 5' and 3' termini of the amplified fragment; a Shine-Delgarno ribosome binding site (RBS) sequence positioned upstream of the initiation codon to ensure efficient translation of the transcribed gene in *E. coli*, and the sequence AACA inserted between positions -4 and -1 for efficient translation in plants.

The PCR-amplified fragment was subcloned into pUC119 for bacterial expression, and

into an episomal yeast plasmid vector, pVT103-U (Vernet *et al.*, 1987 Gene *52*, 225-233) which contains a yeast 2μM origin of replication and a truncated form of the yeast ADH1 promoter to allow high level expression of the cloned GFP gene in *Saccharomyces cerevisiae*. *S. cerevisiae* MGLD-4a (a, *leu2*, *ura3*, *his3*, *trp1*, *lys2*) cells were transformed using the lithium acetate method described by Ito *et al.* (1983).

Transformed *E. coli* and yeast bearing the recombinant plasmid were observed to produce brilliantly fluorescent colonies under long wavelength UV illumination using a hand-held lamp. Interestingly, a high degree of sectoring was seen in yeast colonies containing the episomal form of the *gfp* cDNA; the sectoring was eliminated by integration of the *gfp* cDNA at the yeast URA3 locus and presumably reflects the instability of the 2μM-based episome and/or some toxic effects of GFP expression. This observation also indicated the utility of GFP as a simple cell-autonomous marker. Examination of transformed yeast cells by confocal microscopy showed that the protein was predominantly distributed throughout the cytoplasm (Fig. 1B).

After the PCR amplified *gfp* cDNA was shown to correctly produce fluorescent protein product in yeast, the sequence was inserted between the *Bam*HI and *Sac*I sites in the plant transformation vector pBI121 behind the 35S promoter (Jefferson *et al.*, 1987). *A. tumefasciens* strain LBA4044 was transformed with the *gfp*-containing plasmid by electroporation. Roots of *Arabidopsis thaliana* C24 were transformed using the protocol of Valvekans *et al.* (1988). Transgenic callus and shoots were screened for GFP expression using an inverted fluorescence miscroscope (Leitz DM-IL) fitted with an appropriate filter set (Leitz-D). However, at no stage during the transformation procedure was GFP-related fluorescence detected by UV lamp illumination, or by epifluorescence microscopy. This lack of fluorescence was unexpected and surprising in view of the fluorescence exhibited previously by the transformed bacteria and yeast cells.

gfp* is mis-spliced in *Arabidopsis

The successful expression of GFP in *Arabidopsis* requires proper production of the apoprotein, before post-translational modification to form the chromophore. The inventors therefore used PCR-based methods to verify the correct insertion of the 35S

promoter-driven *gfp* cDNA, and to check mRNA transcription and processing in transformed plantlets. Nucleic acids were extracted from plantlets and either treated with RNase, or DNase treated and reverse transcribed using oligo(dT)$_8$ primer. The *gfp* sequences in these extracts were therefore derived from genomic DNA or transcribed mRNAs, respectively. The *gfp* sequence was PCR-amplified from these separate extracts and products were analysed by restriction endonuclease digestion, as shown in Figure 2A.

In Figure 2A, the RNase-treated DNA sample ("mRNA") is shown on the left of each pair of samples, whilst the DNase-treated/reverse transcribed sample is shown on the right of each pair. The samples were either loaded onto gel without prior restriction ("uncut", extreme left hand pair of samples) or loaded after prior digestion with (from left to right): *Nco*I; *Rsa*I; *Dra*I; *Acc*I; *Hinc*II; or *Ava*II. It can be seen that whilst the expected product was obtained after amplification of the gene, RT-PCR of mRNA sequences gave rise to a truncated product. This product was 80-90 base pairs shorter than expected and was uncut by the restriction endonucleases *Dra*I and *Acc*I, whilst the gene sequences contained unique recognition sites for these enzymes. The inventors established that this is consistent with a small deletion within the *gfp* coding sequence as shown in Figure 2B. In this figure, the shaded portion represents that missing from the mRNA-derived RT-PCR amplified sequences.

The shortened RT-PCR product was cloned and sequenced (Fig. 3A), and a deletion of 84 nucleotides between residues 400-483 was located. The nucleotide sequences bordering this deletion are shown in figure 3B, and demonstrate similarity to known plant introns. The sequence across the splice site (marked with an arrow in Figure 3A) thus reads (5' to 3') ...AG/AC... . Matches were found for important residues at the 5' and 3' splice sites (reviewed by Luefrsen *et al.*, 1994) and the excised *gfp* sequence contains a high predicted A:U content (68%) which has also been shown to be important for recognition of plant introns (Wiebauer *et al.*, 1988; Hanley & Schuler, 1988; Goodall & Filipowitz, 1989, 1991). It is therefore likely that this 84 nucleotide region of the jellyfish *gfp* cDNA sequence is recognised as an intron when transcribed in *Arabidopsis*, resulting in the production of a defective protein product. It should be noted that the borders of this cryptic intron do not coincide with any of the natural spliced junctions

found after processing of the *gfp* mRNA in *A. victoria*.

Modification of the *gfp* gene

The jellyfish gene was mutated to produce a modified *gfp* (*m-gfp*) suitable for expression in *Arabidopsis*, as described below.

Two mutagenic oligonucleotides were synthesized, a 122-mer:

GATCATATGAAGCGGCACGACTTCTTCAAGAGCGCCATGCCTGAGGGATACG
TGCAGGAGAGGACCATCTTCTTCAAGGACGACGGGAACTACAAGACACGTG
CTGAAGTCAAGTTTGAGGG (Seq. ID No. 3),

and a 126-mer:

GATGTATACGTTGTGGGAGTTGTAGTTGTATTCCAACTTGTGGCCGAGGATG
TTTCCGTCCTCCTTGAAATCGATTCCCTTAAGCTCGATCCTGTTGACGAGGGT
GTCTCCCTCAAACTTGACTTC (Seq. ID No. 4).

The oligonucleotides were purified by electrophoresis in a 5% polyacrylamide gel containing TBE and 7M urea. The gel was stained briefly with 0.05% toluidine blue, and the full-length oligonucleotides were excised, and eluted overnight in 0.5M ammonium acetate, 0.1mM Na_2EDTA, 0.1% SDS. The oligonucleotides share 17 nucleotides of complementarity at their 3' termini, and were annealed and elongated after several rounds of thermal cycling with Vent polymerase. The extended product was cloned between the *Nde* I and *Acc* I sites of *gfp*. The mutant clones were screened for the presence of the diagnostic restriction endonuclease sites, *Cla* I, *Ava* II, and the desired fragment (*m-gfp*) was subcloned into M13 and its sequence verified by DNA sequencing using the dideoxynucleotide chain termination technique with T7 DNA polymerase.

The modifications introduced by the synthetic oligonucleotides were intended to alter the sequences which might be involved in 5' splice site recognition and to decrease the A:U content of the putative intron, as shown in Figure 4. In the Figure, the upper DNA sequence is that of a portion of the wild type *A. victoria* GFP. The lower DNA sequence

is that of a portion of a modified GFP-coding sequence. The corresponding amino acid sequence is shown beneath the DNA sequences. Modified nucleotides are shown outlined. All DNA modifications affect only codon usage, and the *m-gfp*-encoded amino acid sequence is identical to that of the wild-type jellyfish polypeptide. The "pseudo-intron" sequence is underlined and the cryptic splice junctions are arrowed. Nucleotide and amino acid residue numbering (on the left and right, respectively, of the oblique stroke) start from the initiation codon. The boxed hexanucleotide sequences are *Nde*I and *Acc*I recognition sites respectively.

The *m-gfp* sequence was inserted behind the 35S promoter in pBI121, and introduced into *Arabidopsis* using the root transformation technique. Brightly green fluorescent cells were seen after co-cultivation with *Agrobacterium*. As shoot regeneration progressed, explants with different levels of green fluorescence could be observed. Regenerating callus and shoots develop a bright red autofluorescence due to the formation of chlorophyll within the tissues, and with the brightest *m-gfp* transformants the green fluorescence was clearly detectable against this autofluorescent background using a hand held UV lamp. This was similar to the levels of green fluorescence seen in transformed yeast and *E. coli*. However, these very bright *Arabidopsis* transformants regenerated and set seed rather poorly. Nevertheless, seeds were obtained from over 50 transformed lines, allowed to germinate, and screened by epifluorescence microscopy. Several of the brightest lines were used for confocal laser scanning microscopy.

Confocal microscopy of living plants

The fluorescence properties of GFP and chlorophyll allow the use of fluorescence microscopes equipped with common filter sets for fluorescein and rhodamine for dual imaging in plant cells. Intact five day old *m-gfp*-transformed *Arabidopsis* seedlings were mounted in water for confocal laser scanning microscopy. GFP fluorescence could be clearly visualised in the transformed tissues, and chloroplasts provided a very effective counter fluor in the upper parts of the plant. Optical sectioning of the *m-gfp* transformed plants gives selective visual access to the internal details of living plant structure, as shown in Figure 5, without any need for staining or dissection. For example, median longitudinal sections of root tips can be simply obtained by adjusting the microscope

depth of focus, and confocal imaging allows the resolution of subcellular details. GFP is found throughout the cytoplasm, but appears to accumulate within the nucleoplasm. It appears excluded from vacuoles, organelles and other bodies in the cytoplasm, and is excluded from the nucleolus. Similarly, in optical sections of cotyledon and hypocotyl tissues, GFP is found throughout the cytoplasm and nucleoplasm. The relationship of cells within the tissues is clearly discernible. In highly vacuolate epidermal cells in the root and hypocotyl, GFP fluorescence allows visualisation of trans-vacuolate cytoplasmic threads, and the thin cytoplasmic strands which underly the cell wall and which may be aligned with cytoskeletal elements. The movement of organelles through cytoplasmic streaming could also be observed in these living cells.

Example 2

Isolation and characterisation of a bright mutant of GFP

The sequence of *m-gfp* was mutated by PCR in the presence of limiting nucleotide concentrations. The template plasmid was pBSm-gfp4, a derivative of TU#65 (Chalfie *et al.*, 1994 Science *263*, 802-805) in which *gfp* has been replaced with *m-gfp*. The primers used were the T3 and T7 primers (New England Biolabs) that are complementary to the flanking T3 and T7 promoters present in the vector sequence. Four separate reactions (30 cycles of 30 seconds at 94°C, 30 seconds at 55°C and 1 min at 72°C using Taq DNA Polymerase from Promega) were carried out, each with the concentration of a different nucleotide reduced from 200 μM to 20 μM. The amplified fragments were pooled, cleaved with *Kpn*I and *Eco*RI and cloned downstream of the *lac* promoter of pBluescript II KS (+) (Stratagene).

The mutant library thus obtained was transformed into *E. coli* strain XL1-Blue (Stratagene) and incubated overnight at 37°C on TYE agar containing 50 μg/ml ampicillin and 1 mM IPTG. Colonies were illuminated with a long wavelength UV lamp (UVP Model B 100 AP) and visually screened for increased fluorescence. The coding regions of two of the brightest mutant genes (*m-gfpA* and *m-gfpB*) thus identified, as well as that of *m-gfp*, were amplified by PCR (30 cycles of 1 min at 94°C, 1 min at 55°C and 1 min at 72°C using VENT DNA Polymerase from New England Biolabs) using primers

that generate a *Bam*HI-*Sac*I fragment containing the *gfp* coding sequence downstream of a phage Shine-Dalgarno sequence and a plant translation initiation context sequence. The forward primer (5'Bam-GFP) was

5'-GGCGGATCC*AAGGA*GATATAACAATGAGTAAAGGAGAAGAACTTTTCACT-3'

(Seq ID No. 5. *Bam*HI site underlined. Shine-Dalgarno sequence in italics and translation initiation context sequence in bold) and the reverse primer (GFP-3'Sac) was 5'-GGCGAGCTCTTATTTGTATAGTTCATCCATGCC-3' (Seq ID No. 6. *Sac*I site underlined and GFP stop codon in bold). The amplified fragments obtained from the three reactions were cleaved with *Bam*HI and *Sac*I and cloned downstream of the *lac* promoter of pUC119 (Vieira and Messing, 1987 Methods Enzymol. *153*, 3-11).

The positions of the mutations responsible for the bright phenotypes of *m-gfpA* and *m-gfpB* were then localised by recombination of the mutant genes with *m-gfp*. The pUC119 derivatives containing *m-gfpA* and *m-gfpB* were cleaved with either *Bam*HI and *Nco*I, *Nco*I and *Cla*I, or *Cla*I and *Sac*I. The restriction fragments were gel purified and ligated to the *m-gfp* pUC119 derivative that had been cleaved with the same combination of enzymes and gel purified. These and the parent constructs were introduced into XL1-Blue cells and incubated overnight at 37°C on agar plates containing ampicillin and IPTG. Comparison of the fluorescence of colonies containing the various constructs revealed that the mutation(s) responsible for the bright phenotypes of both *m-gfpA* and *m-gfpB* were contained within the 336 bp *Cla*I-*Sac*I fragment at the 3' end of the gene. These fragments were cloned into the phage vector M13mp19 (New England Biolabs) and sequenced from the Universal primer using the Sequenase Version 2.0 DNA Sequencing Kit (United States Biochemical Corporation).

Sequencing of the *Cla*I-*Sac*I fragment of *m-gfpB* revealed the presence of a single coding alteration, V163A. This same change was found in combination with a second coding alteration, S175G, in the *Cla*I-*Sac*I fragment of *m-gfpA*. The sequences of mGFP, mGFPB and mGFPA are compared in Figure 6, which shows the location of the two mutations. The S175G change would appear to contribute to the phenotype of GFPA as cells expressing the GFPA protein were clearly more fluorescent than those expressing GFPB (data not shown). Thus, only GFPA was analysed further.

The sequence of the *m-gfp* gene was modified so as to code for the V163A and S175G substitutions described above and the I167T substitution described in the prior art (1994 Heim et al., Proc. Natl. Acad. Sci. USA *91*, 12,501-12,504, which substitution inverts the ratio of the 400-475nm excitation peaks), as well as to further alter the codon usage of the gene in order to eliminate potential plant intron sequences generated by the introduction of these mutations. The sequence differences between this modified gene termed, *m-gfp5*, and the original *gfp* gene, and their respective polypeptides are summarised in Fig. 15 (Seq. ID Nos. 5-8).

The *m-gfp5* gene was constructed by PCR amplification (30 cycles of 30 secs at 94°C, 30 secs at 55°C and 30 secs at 72°C using VENT DNA Polymerase) of *m-gfp* using mutagenic primers. The forward primer was an oligo corresponding to nucleotides 445-560 of the *m-gfp5* coding sequence shown in Figure 15 and the reverse primer was GFP-3'Sac. The amplified fragment was cleaved and exchanged with the *Acc*I-*Sac*I fragment of *m-gfp* to create *m-gfp5*.

For bacterial expression studies, *Bam*HI-*Sac*I PCR fragments containing the *m-gfp*, *m-gfpA* and *m-gfp5* genes were cloned downstream of the *tac* promoter of the expression vector pSE380 (Invitrogen), to give the plasmids pSE-GFP, pSE-GFPA and pSE-GFP5, respectively. Expression from the *tac* promoter of pSE380 is tightly regulated due to the presence on the plasmid of the *lac*Iq gene. For yeast expression, the same PCR fragments containing the *m-gfp*, *m-gfpA* and *m-gfp5* genes were inserted downstream of the constitutive *ADH1* promoter of pVT103-U (Vernet *et al.*, 1987), a yeast multicopy episomal plasmid containing the *URA3* selectable marker. The resulting plasmids were pVT-GFP, pVT-GFPA and pVT-GFP5, respectively.

To assess the difference in fluorescence between strains expressing modified GFP and GFPA quantitatively, the inventors introduced expression plasmids pSE-GFP and pSE-GFPA into *E. coli* strain XL1-Blue and measured the fluorescence (λ_{ex} = 397 nm, λ_{em} = 508 nm) of equal optical densities of cells at various times following IPTG-induction of protein synthesis at 37°C (Fig. 7). 4.5 hrs after induction, cells expressing m-GFPA were observed to fluoresce approximately 20-fold more intensively than those expressing

m-GFP, a figure which increased to approximately 35-fold by the time the cells had entered stationary phase (9 hrs after induction).

To determine whether the enhanced fluorescence of cells expressing m-GFPA might be due to increased levels of protein expression, total protein was prepared from cells from the 4.5 hr time point and the amount of intracellular GFP estimated by Western blot analysis. As can be seen in Fig. 8, m-GFPA accumulates to a significantly higher level than modified GFP. The "vector" track is a negative control. The numbers on the right of the blot represent the molecular weights of known standards. However, the difference in protein levels as estimated by quantification of band intensities, 2.4-fold, is not nearly enough to account for the approximately 20-fold difference in fluorescence levels observed at this time point. This result suggests that a large proportion of GFP that is expressed in cells at 37°C is non-fluorescent and that the substitutions present in m-GFPA enhance the maturation of the protein to the fluorescent form. Comparison of the growth curves of strains expressing m-GFP and m-GFPA with the growth curve of a non-expressing strain (data not shown) indicated that expression of these proteins does not have any adverse effects on the growth of bacterial cells.

The amino acid substitutions present in m-GFPA suppress the temperature-sensitivity of GFP maturation

Lim and co-workers (Lim *et al.*, cited above) have recently reported that maturation of GFP to the fluorescent form is sensitive to temperature during expression in the yeast *Saccharomyces cerevisiae*. To test whether the same may be true during expression in *E. coli* and whether the substitutions present in m-GFPA enhance maturation by suppressing any such sensitivity, the inventors examined expression of m-GFP and m-GFPA over a range of different temperatures. Strains containing pSE-GFP and pSE-GFPA were induced overnight at temperatures ranging between 25°C and 42°C. For each culture, the fluorescence of equal optical densities of cells was measured and the amount of intracellular GFP determined by Western blot analysis (Table 1). Fluorescence values were then normalised against the amount of GFP present inside cells so as to give a relative measure of the proportion of intracellular GFP that is fluorescent for each culture. The results (Fig. 9) clearly show that the proportion of modified GFP

that is fluorescent steadily decreases with increasing incubation temperature (open columns), indicating that maturation of the protein to the fluorescent form is temperature sensitive. In contrast, the substitutions present in m-GFPA (shaded columns) suppress this sensitivity to temperature, with maturation being optimal at 37°C.

Table 1

Temperature (°C)	Fluorescence (arbitary units)		Relative amount of intracellular GFP (units)	
	m-GFP	m-GFPA	m-GFP	m-GFPA
25	328.4	722.2	0.29	0.58
30	100.5	541.1	0.21	0.82
37	67.9	2273.0	0.23	1.00 ($7.8x10^5$)
42	9.2	369.4	0.17	0.44

Investigation into the thermosensitivity of GFP maturation

The post-translational maturation of GFP to the fluorescent form involves a number of steps. The first step, presumably, is folding of the apoprotein into a catalytic conformation that facilitates the novel reactions involved in formation of the chromophore. These reactions consist of cyclisation and oxidation of the tripeptide Ser65-Tyr66-Gly67 to give a p-hydroxybenzylidene-imidazolidinone structure. Once the chromophore has been formed, it is then only fluorescent once GFP has adopted a fold which protects it from solvent effects. In principle, any of these processes could be sensitive to temperature and thus be responsible for the observed thermosensitivity of GFP maturation.

Since the oxidation reaction involved in chromophore formation appears to require molecular oxygen, Heim and co-workers (Heim *et al.*, 1995 Nature *373*, 663-664) have been able to measure the reaction rate by expressing GFP in *E. coli* under anaerobic conditions and then monitoring the development of fluorescence after admission of air.

To determine whether this reaction might be temperature sensitive and whether the substitutions present in m-GFPA act by enhancing its rate at higher temperatures, we measured the rates of oxidation of m-GFP and m-GFPA at both 25°C and 37°C. For these experiments a yeast expression system was used, which provided for better growth and expression levels than *E. coli* under anaerobic conditions. Strains of *Saccharomyces cerevisiae* MGLD-4a (*MATα, leu2, ura3, his3, trp1, lys2*) containing either pVT-GFP or pVT-GFPA were incubated anaerobically (Becton-Dickinson BBL GasPak Pouch) overnight at 30°C in synthetic drop-out media lacking uracil (Rose *et al.*, 1990 "Methods in Yeast Genetics. A Laboratory Course Manual", Cold Spring Harbor Laboratory Press, Cold Spring Harbor, USA). Following admission of air to the pouch, 1.0 ml of each culture was immediately centrifuged for 1 min at 13,000 rpm and resuspended in 0.5 ml aerated and prewarmed PBS (pH 7.4) containing 8 mM NaN_3 as a metabolic inhibitor. Cell suspensions were placed immediately into pre-warmed cuvettes held within the fluorimeter carousel and the time course of fluorescence (λ_{ex} = 397 nm, λ_{em} = 508 nm) development measured.

As reported previously by Heim *et al.*, each oxidation proceeded as a simple first order reaction (Fig. 10). Figure 10 shows the rate of fluorescence development for cultures expressing modified GFP at 25°C (crosses) or 37°C (triangles), or cultures expressing GFPA at 25°C (squares) or 37°C (circles).

The time constant measured for the oxidation of m-GFP at 37°C (5.9 $\pm$ 0.1 min) was found to be approximately 3-fold faster than that measured at 25°C (16.2 $\pm$ 0.3 min), indicating that the post-translational oxidation of the GFP chromophore is not the step responsible for the temperature sensitivity of maturation. In confirmation of this conclusion, the time constants derived for m-GFPA at both 25°C and 37°C (22.5 $\pm$ 1.4 min and 18.1 $\pm$ 0.4 min, respectively) were actually slower than those measured for m-GFP.

Heim and co-workers have also reported that some GFP forms non-fluorescent inclusion bodies during expression in *E. coli*, indicating that not all GFP folds properly under these conditions. To determine whether the proper folding of m-GFP might be temperature

sensitive and whether the substitutions present in m-GFPA act by enhancing proper folding at increased temperatures, the inventors examined the solubilities of the two proteins during expression in *E. coli* at 25°C and 37°C. Bacterial cells expressing m-GFP or m-GFPA were grown overnight at either 25°C or 37°C, lysed, and the soluble and insoluble fractions separated by centrifugation.

Specifically, cells containing pSE-GFP or pSE-GFPA were grown in 1.5 ml of 2xTY broth to an absorbance of 0.2 at 600 nm and then induced overnight with 0.2 mM IPTG. The cultures were centrifuged at 13,000 rpm for 2 min, resuspended in 500 μl 50 mM Tris-HCl (pH 8.0), 2 mM EDTA, 100 μg/ml lysozyme, 0.1% Triton X-100 and incubated at 30°C for 15 min. Cells were then lysed by sonication (5 x 15 secs) using a Heat Systems (Model CL4) sonicator and centrifuged at 13,000 rpm for 15 min at 4°C. The supernatant (soluble fraction) was removed and stored at -70°C until used. The pellet (insoluble fraction) was washed once with 500 μl 50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 0.5% Triton X-100, resolubilised for 1 hr at room temperature in 500 μl resolubilisation buffer (8 M urea, 0.1 M NaH_2PO_4, 10mM Tris-HCl, pH 6.3) and stored at -70°C until used. The amount of m-GFP or m-GFPA present in each fraction was then estimated by Western blot analysis (Fig. 11).

SDS-PAGE and Western blot analysis were carried out according to normal procedure (Sambrook *et al.*, 1989 "Molecular Cloning. A Laboratory Manual". Cold Spring Harbor Laboratory Press, Cold Spring Harbor, USA). Primary antibodies were polyclonal rabbit anti-GFP (generous gift of S. Santa-Cruz) used at a dilution of 1/2,000. Antibodies were detected with iodinated Protein A (Amersham) and bands visualised and quantified using a Molecular Dynamics Phosphorimager. In Figure 11, insoluble and soluble fractions are denoted by the letters I and S respectively, with cultures grown at 25°C shown on the left, and cultures grown at 37°C shown on the right. In all cases, fluorescence was found almost exclusively in the soluble fraction. At 25°C, both m-GFP and m-GFPA were found predominantly in the soluble fraction, indicating that proper folding of both proteins is efficient at this temperature. At 37°C, however, the majority of m-GFP was found as aggregated protein in the insoluble fraction, whereas most of m-GFPA was still present in the soluble fraction. This result indicates that the temperature

sensitivity of m-GFP maturation is due primarily to improper protein folding at higher temperatures and that this defect is suppressed by the amino acid substitutions present in m-GFPA.

To gain information on which species in the maturation pathway of GFP aggregates at higher temperatures, the inventors made use of the characteristic absorption of the GFP chromophore in either the mature (Ward & Bokman, 1982 Biochemistry *21*, 4535-4540) or chemically reduced state (Inouye & Tsuji 1994 FEBS Lett. *351*, 211-214). If the aggregating species has already undergone the cyclisation reaction, GFP isolated from inclusion bodies should show this characteristic absorption. To facilitate the purification of protein for absorbance measurements, the inventors fused a polyhistidine tag to the C-terminus of m-GFP.

Histidine-tagging was achieved by the addition of 6 histidine codons to the 3' ends of the modified *gfp* genes by PCR. The genes were amplified using 5'Bam-GFP as the forward primer and the oligo

5'- GCCGAGCTCTTA**GTGGTGGTGGTGGTG**

TTTGTATAGTTCATCCATGCC -3'

(Seq ID No. 7. *Sac*I site underlined, histidine codons in bold) as the reverse primer. The amplified fragments were cleaved with *Bam*HI and *Sac*I and cloned downstream of the *tac* promoter of pSE380 (Invitrogen) to give the expression plasmids pSE-GFPHis, pSE-GFPAHis and pSE-GFP5His respectively.

For the purification of histidine-tagged GFP for absorbance measurements, soluble and insoluble fractions of cells containing pSE-GFPHis grown at 25°C and 37°C, respectively, were prepared as described previously. GFP was purified from the fractions on Ni-chelate columns using the Ni-NTA Spin Kit (Qiagen). Purification from the soluble fraction was carried out according to the protocol for the purification of histidine-tagged proteins under native conditions. After clearance of cellular debris from the insoluble fraction by centrifugation at 13,000 rpm for 30 min, purification was carried out according to the protocol for purification of histidine-tagged proteins under denaturing conditions, except that the protein was eluted with resolubilisation buffer containing 250

mM imidazole.

For the purification of histidine-tagged proteins for fluorescence spectroscopy, cells were grown in 100 ml of 2xTY broth at 37°C to an absorbance of 0.2 at 600 nm and then induced overnight with 0.5 mM IPTG. Cells were harvested by centrifugation at 6,000 rpm for 10 min and lysed by resuspension in 4 ml 20 mM Tris-HCl (pH 7.9), 500 mM NaCl, 5 mM imidazole, 0.1% sarkosyl, 0.1% deoxycholate, 2.25 M Guanidine-HCl. Nucleic acids were precipitated by the addition of 5ml isopropanol and removed by centrifugation at 10,000 rpm for 10 min. Fluorescent histidine-tagged proteins were purified from the supernatant on Ni-chelate columns (Qiagen) and eluted with 2ml of 20mM Tris-HCl (pH 7.9), 500 mM NaCl, 150 mM imidazole. For all purifications, protein purity was assayed by SDS-PAGE and found to be >95%. Protein concentrations were determined by Bradford assay (Bio-Rad Protein Assay kit) using bovine serum albumin as a standard.

Absorbance spectra were recorded on a Cary 3 UV-Visible Spectrophotometer (Varian) at 25°C. The optical pathlength was 1 cm. Fluorescence spectra were recorded on a Hitachi F-4500 fluorimeter at 25°C using 4mm/10mm cuvettes. The bandpass for both the excitation and the emission monochromators was 5 nm, the scan speed 240 nm per min and the response time automatically adapted by the device. All spectra were corrected following the supplier's procedure for calibration of the fluorimeter using Rhodamine-B as standard. Emission spectra were recorded at a fixed wavelength of the excitation maximum, excitation spectra at a fixed wavelength of the emission maximum.

Histidine-tagging of GFP did not detectably affect the temperature sensitivity of maturation of the protein (data not shown). The absorption spectra of equal concentrations of denatured protein from the two preparations were then recorded, as described above. The results are shown in Figure 12, which is a graph of absorbance against wavelength (nm).

As can be seen in Fig. 12, denatured fluorescent protein derived from the soluble fraction of cells grown at 25°C shows a characteristic absorption peak similar to that of the m-

GFP chromophore at acid pH (Ward & Bokman, cited above). On the other hand, protein purified from inclusion bodies of cells grown at 37°C shows no such absorption, indicating that the aggregating species has not formed a chromophore. Taken together, the results presented above indicate that the temperature sensitivity of m-GFP maturation is due primarily to the failure of the unmodified apoprotein to fold into its catalytically active conformation at higher temperatures. Furthermore, the amino acid substitutions present in m-GFPA suppress this defect by enhancing proper folding at elevated temperatures.

If thermosensitivity of m-GFP maturation observed in the yeast *Saccharomyces cerevisiae* is also a result of the thermosensitivity of apoprotein folding, it should be suppressed by the substitutions present in m-GFPA. To test this prediction, the inventors incubated strains of *cerevisiae* containing either pVT-GFP or pVT-GFPA on agar plates at either 25°C or 37°C. As can be seen in Fig. 13, the substitutions present in m-GFPA also suppress the thermosensitivity of m-GFP expression in yeast. This result indicates that the temperature-dependent mis-folding of the m-GFP apoprotein is not simply an artefact of an *E. coli* overexpression system, but is also the basis for the thermosensitivity of m-GFP maturation in a heterologous eukaryotic system.

Modification of the fluorescence spectra of m-GFPA

Fluorescence spectroscopy of purified histidine-tagged m-GFP and m-GFPA revealed that the fluorescence spectra of m-GFPA are essentially unchanged from those of m-GFP except for a decrease in the amplitude of the 475 nm excitation peak relative to the amplitude of the 400 nm excitation peak (Fig. 14). Although this spectral change is advantageous for applications which utilise 400 nm excitation, it is also detrimental for those which utilise 475 nm excitation. For many experiments, the ideal spectral variant would be a protein which could be efficiently excited at either of these wavelengths. This characteristic would afford greater flexibility with regard to the range of applications in which the protein could be used.

Recently, it has been demonstrated that, as observed here, the relative amplitudes of the excitation peaks of GFP can be altered by means of mutagenesis (Ehrig *et al.*, 1995

FEBS Lett. *367*, 163-166; Delagrave *et al.*, 1995 Bio/Technology *13*, 151-154). A number of these mutations, like the substitutions present in m-GFPA, are located in the C-terminal region of the protein. It has been hypothesised that, in the three-dimensional structure of GFP, the C-terminal region is close to the chromophore and that mutations in this region can affect the microenvironment of the chromophore so as to influence the equilibrium between the two tautometric forms of the chromophore that are responsible for the two excitation peaks (Heim *et al.*, 1994 & Ehrig *et al.*, 1995). One of these mutations, I167T, inverts the ratio of the 400 nm to 475 nm excitation peak heights. If the effects of mutations in the C-terminal region of GFP on the spectroscopic state of the chromophore are additive, then it is possible that combination of the I167T substitution with the substitutions present in GFPA might increase the amplitude of the 475 nm peak relative to the 400 nm peak.

Histidine-tagged m-GFP5 was purified and its excitation and emission spectra analysed by fluorescence spectroscopy. As can be seen in Fig. 14, m-GFP5 has two excitation peaks (maxima at 395 nm and 473 nm) of almost exactly equal amplitude and an emission spectrum largely unchanged from that of m-GFP. To determine whether m-GFP5 has retained the thermotolerant phenotype of GFPA, bacterial cells containing pSE-GFP or an expression plasmid containing *m-gfp5* (pSE-GFP5) were induced with IPTG for 5 hours at 37°C. The fluorescence (λ_{ex} = 395 nm or 473 nm, λ_{em} = 507 nm) of equal optical densities of cells was then measured. Cells expressing m-GFP5 were observed to fluoresce 39-fold more intensely than cells expressing m-GFP when excited at 395 nm and 111-fold more intensely when excited at 473 nm. These results indicate that m-GFP5 has not only retained the thermotolerant phenotype of m-GFPA, but has improved upon it.

Further modification of *mgfp5* was achieved. Two synthetic oligonucleotides were made, to act as mutagenic PCR primers to add an in-frame *Eco*RI site at the 5' end of the gene and to add a sequence coding for the amino acid tag HDEL at the C terminal of the protein (which tag acts as an endoplasmic reticulum localisation signal).

The PCR mutagenised sequence was then used in a three-way ligation reaction with

*Bam*HI/*Sac*I - cut vector and a pair of synthetic oligonucleotides with *Bam*HI/*Eco*RI ends. The synthetic oligos had the sequences:

5' GGC GGA TCC AAG GAG ATA TAA CAA TGA AGA CTA
ATC TTT TTC TCT TTC TCA TCT TTT CAC 3' (Seq ID No. 8) and

5' GCC GAA TTC GGC CGA GGA TAA TGA TAG
GAG AAG TGA AAA GAT GAG AAA GA 3' (Seq ID No. 9)

The oligos were annealed, extended with Klenow polymerase and cut with *Bam*HI/*Eco*RI. When ligated with the *m-gfp5*HDEL gene, they introduced the signal sequence from *Arabidopsis* chitinase at the 5' end of the coding sequence. The nucleotide sequence of the resulting modified gene (*m-gfp5-ER*), and the amino acid sequence of its polypeptide product (Seq ID No.s 10 and 11 respectively), are shown in Figure 16. The nucleotides encoding the signal sequence are shown in upper case letters, whiolst the rest of the sequence is in lower case letters. The C terminal HDEL tag on the protein is apparent.

The modified gene, when expressed in Arabidopsis, gave highly efficient concentration of GFP-mediated fluorescence in the endoplasmic reticulum (Figure 17). Referring to Figure 17, the panels illustrate confocal micrographs of 5-day old *A. thaliana* seedlings expressing m-GFP (panels A-D) or m-GFP5-ER (panels E-H), imaged at 395nm excitation wavelength. Panels A & E are sections of the junction between the hypocotyl and the cotyledon (bar = 25μM); panels B & F show hypocotyl epidermal cells (bar = 10μM); panels C & G show median longitudinal sections of root tips (bar = 25μM); and panels D & H show roots, with nucleoplasmic accumulation of mGFP (in D) and retention in the endoplasmic reticulum of mGFP5-ER (in H), (bar = 10μM).

DISCUSSION

GFP expression in plants

The objective of this work was to begin the development of *gfp* for use as a genetic marker in transformation and as a reporter for localised gene expression in *Arabidopsis*

and other plants. In order to successfully employ the *gfp* cDNA in plants, three major steps need to be addressed.

(1) The GFP apoprotein must be produced in suitable amounts within the plant cells.

(2) The apoprotein must undergo efficient post-translational cyclisation and oxidation to produce the mature GFP.

(3) The fluorescent protein may need to be suitably targeted within the cell, to allow efficient post-translational processing, safe accumulation to high levels, or to allow easier distinction of expressing cells.

The inventors have shown that expression of the jellyfish *gfp* cDNA in *Arabidopsis* is curtailed by aberrant splicing, with an 84 nucleotide intron being efficiently excised from within the GFP coding sequence. The recognition of introns in plant pre-mRNAs primarily requires conserved sequences found adjacent to the 5' and 3' splice sites, which are related to those found in other eukaryotes, and, atypically, a high A:U content within the intron. The inventors altered potential recognition sequences at the 5' splice site, and decreased the A:U content of the cryptic intron by *in vitro* mutagenesis to produce a modified *m-gfp* gene which was successfully expressed in transgenic *Arabidopsis* plants. It is likely that this *m-gfp* gene will be useful for expression studies in other plants, which appear to share similar features involved in intron recognition.

It is also possible that aberrant splicing may interfere with GFP expression in other organisms. However, introns found in yeast possess a requirement for conserved sequences located at the branch point, and introns found in animal cells (including jellyfish) share a conserved polypyrimidine tract adjacent to the 3' splice site. The lack of these additional features may allow correct processing of the *gfp* mRNA in fungal and animal cells.

With expression of the *m-gfp* gene in *Arabidopsis*, it has now been shown that the apoGFP readily undergoes maturation, and that the fluorescent form of the protein accumulates in transformed cells. Transformed cells were often intensely fluorescent, and were easily detectable by eye using a long-wave UV lamp. However it proved difficult to efficiently regenerate fertile plants from the brightest transformants, with cells

remaining as a highly fluorescent callus or mass of shoots after several months of culture. It is possible that high levels of GFP expression were mildly toxic or interfered with differentiation, due perhaps to the fluorescent or autocatalytic properties of the protein. In the natural situation, in jellyfish photocytes where high levels of GFP are well tolerated, the protein is found sequestered in microbody-like lumisomes. In contrast, the mature protein is found throughout the cytoplasm and nucleoplasm in transformed *Arabidopsis*. If GFP is a source of fluorescence-related free radicals, for example, it might be advisable to target the protein to a more localised compartment within the plant cell. Appropriate localisation signals are known to those skillled in the art and it should prove possible to incorporate these into the GFP polypeptide without unduly disrupting the fluorescence characteristics of the protein.

The inventors have adapted the green fluorescent protein (GFP) of *Aequoria victoria* for use as a genetic marker in *Arabidopsis thaliana*. Transcripts of the jellyfish GFP coding sequence are mis-spliced in *Arabidopsis*, with an 84 nucleotide intron being efficiently excised. A modified version of the *gfp* sequence has been constructed to destroy this cryptic intron, and to restore proper expression of the protein in plant cells. GFP is mainly localised within the nucleoplasm and cytoplasm within transformed *Arabidopsis* cells, and its presence allows optical sectioning of intact plants using confocal laser scanning microscopy. The modified *gfp* sequence may be useful for directly monitoring gene expression and protein localization at high resolution, and as a simply scored genetic marker in living plants.

A major use for *m-gfp* would be as a replacement for the β-glucuronidase gene, used as a reporter for promoter and gene fusions in transformed plants. Histochemical staining is used to identify cells expressing the GUS gene product, but a fluorescent product can be imaged directly and rapidly. Gene expression and protein localization can be observed in physiologically active cells without a prolonged and lethal staining procedure, and fluorescence microscopy techniques allow the high resolution imaging of GFP-expressing cells. In addition, it becomes feasible to follow dynamic events in living cells and tissues.

High levels of fluorescence intensity are obtained in GFP-transformed bacterial and yeast colonies allowing simple screening for GFP expression with the use of a hand-held UV lamp. Such an assay for gene expression in living plants would be a very useful tool for plant transformation experiments. Many transformation techniques give rise to regenerating tissues which are variable or chimeric, and require testing of the progeny of the primary transformants. Potentially, *m-gfp*-transformed tissues could be monitored using *in vivo* fluorescence, avoiding any need for destructive testing, and the appropriate transformants could be rescued and directly grown to seed. Similarly, *in vivo* fluorescence would be an easily scored marker for field testing in plant breeding, allowing *m-gfp*-linked transgenes to be easily followed.

Use of a confocal laser scanning microscope will allow the clear analysis of plant tissue whole-mounts despite the refractile nature and light scattering (and for some cells, autofluorescent) properties of plant cell walls.

In this study, the inventors have also shown that maturation of GFP in *E. coli* is sensitive to temperature, due primarily to the mis-folding of the apoprotein into inclusion bodies at elevated temperatures. They have also described two mutants, m-GFPA and m-GFP5, whose folding is thermotolerant. Presumably, the charateristic of the GFP apoprotein that causes it to aggregate at higher temperatures and the mechanism by which the mutations present in m-GFPA and m-GFP5 suppress this effect is unknown. However, studies on the effects of mutations on the tendency of some other proteins to aggregate (Thomas *et al.*, 1995 Trends Biochem. Sci. *20*, 456-459; Mitraki *et al.*, 1991 Science *253*, 54-58; Wetzel 1994 Trends Biotech. *12*, 193-198; and Chrunyk *et al.*, 1993 J. Biol. Chem. *268*, 18053-18061) suggest a number of possibilities.

The simplest explanation is that the native apoprotein or one of its folding intermediates is thermodynamically unstable and the protein aggregates when in the unfolded state. The substitutions present in the thermotolerant mutants could suppress the characteristic either by increasing the thermodynamic stability of the unstable species or by decreasing its steady state level by increasing the rate of chromophore cyclisation. Alternatively, higher temperatures can allow proteins to overcome the thermodynamic barriers to the

formation of off-pathway folding intermediates which may become kinetically trapped by aggregation. It is possible, therefore, that the substitutions present in the heat-tolerant mutants act by suppressing such a phenomenon by directing folding along the correct pathway at elevated temperatures. It is also possible that the kinetic half-life of an aggregation-prone intermediate in the normal apoprotein folding pathway is significantly increased at higher temperatures. Mutations could suppress this characteristic either by decreasing the half life of such an intermediate or by reducing its tendency to aggregate. Finally, the suppressor mutations may favour proper folding at higher temperatures by increasing the affinity of the apoprotein for a molecular chaperone.

To differentiate between these different possibilities, biophysical analyses of variants of GFP and the heat tolerant mutants that cannot undergo cyclisation and are thus trapped as apoproteins will be required. However, the observation that the substitutions present in m-GFPA increase the T_m of mature GFP by 4.0°C (data not shown) provides preliminary evidence that they may act by increasing the thermodynamic stability of the native apoprotein. As well as enhancing proper folding, the substitutions present in m-GFPA contribute to the bright phenotype of the mutant protein by facilitating its accumulation to higher levels than m-GFP (Table 1). This observation indicates that GFP is turned over more rapidly than m-GFPA, probably because partially or mis-folded m-GFP apoprotein that does not aggregate would be degraded by the cellular proteolytic machinery.

The inventors have shown that oxidation of the GFP chromophore does not contribute to the temperature sensitivity of maturation by measuring the reaction rate in yeast cells at both 25°C and 37°C (Fig. 3). An interesting point arising from this experiment is that the time constants derived for m-GFP at both 25°C and 37°C (5.9 $\pm$ 0.1 min and 16.2 $\pm$ 0.3 min, respectively) are significantly faster than the 120 min estimated for the oxidation of GFP in bacteria by Heim *et al.* This observation may reflect a difference in the physiological states of yeast and bacterial cells following anaerobic growth or perhaps the presence of a catalysing factor in yeast cells.

Nevertheless, these results suggest that the oxidation of the GFP chromophore has the

capacity to proceed at a much higher rate than previously thought. Therefore, in some cases, the factor which limits how quickly fluorescent can be observed following protein synthesis may be the efficiency with which the apoprotein folds rather than the time taken for oxidation of the chromophore.

Examination of the fluorescence spectra of the m-GFPA revealed a decrease in the amplitude of the 475nm excitation peak relative to the amplitude of the 400nm excitation peak. This result indicates that mutations in the C-terminal region of GFP are able to modulate the spectroscopic state of the chromophore by affecting its local environment within the protein. The inventors have utilised this phenomenon to engineer the fluorescence spectra of m-GFP5 by introducing a third substitution, I167T, into the C-terminal region of m-GFPA. m-GFP5 has two excitation peaks (maxima at 395nm and 473nm) of almost exactly equal amplitude and is thus ideal as a multi-purpose spectral variant which can be used for applications requiring both UV and blue excitation. Since the substitutions present in m-GFPA and m-GFP5 affect the environment of the chromophore, it is likely that they also influence the intrinsic brightness of the mutant proteins by affecting the extinction coefficients and/or quantum yields of the chromophore at the two excitation wavelengths. However, to measure an extinction coefficient accurately, one must be certain that every GFP molecule in a given sample is mature. The results presented here suggest that, even in a soluble fraction, there may be appreciable amounts of mis-folded or non-fluorescent apoprotein. In support of this notion, the ratio of the absorbance of the chromophore to that of the aromatic amino acids of histidine-tagged m-GFP purified from the soluble fraction of bacterial cells grown at either 25°C (Fig. X+4) or 37°C (Inouye & Tsuji, cited above) is approximately 0.4. Since this value is in excess of 1.0 for either native or acid-denatured GFP isolated directly from the *Aequorea* jellyfish, it would appear that more than half of the recombinant GFP in a soluble fraction does not have a chromophore.

These observations suggest that extinction coefficients and quantum yields may be difficult to measure unambiguously for recombinant forms of GFP. Therefore, great care must be taken when interpreting the effects of mutations that alter the brightness of GFP. For example, a number of mutations in and near the chromophore of GFP have recently

been described that cause significant shifts in the excitation and/or emission spectra of the protein. A subset of these mutations that alter the tyrosine residue at position 66 to tryptophan, histidine or phenylalanine progressively blue-shift the excitation and emission spectra. However, these mutant proteins are much less fluorescent than GFP, a phenomenon which has been attributed to them having sub-optimal extinction coefficients and/or quantum yields due to the poor fit of the alternative amino acids into the central cavity normally occupied by the tyrosine residue. It is possible, however, that the observed low fluorescence of these mutants is due to detrimental effects of the substitutions on folding and/or chromophore formation, resulting in the presence of large amounts of non-fluorescent protein in soluble fractions. Therefore, it is feasible that the proper maturation of these mutants might be enhanced by the introduction of the amino acid substitutions present in m-GFPA or m-GFP5. Indeed, expression of a protein containing the Y66H mutation in combination with the substitutions present in m-GFPA in *E. coli* at 37°C resulted in a 29-fold increase in fluorescence. The fluorescence spectra of this hybrid protein were also unchanged from those of the Y66H mutant alone (data not shown). Therefore, it is foreseeable that the substitutions present in m-GFPA and m-GFP5 may be combined with these and other pre-existing spectral mutations in the chromophore of GFP to produce a range of spectral variants with greatly improved maturation characteristics.

As well as in *E. coli*, maturation of m-GFP appears to be thermosensitive in the yeast *Saccharomyces cerevisiae* and in mammalian cells. Therefore, it appears that the sensitivity of apoprotein folding to temperature may be a ubiquitous phenomenon. Indeed, it is interesting to note that the brightness of m-GFP in *Arabidopsis thaliana* is markedly increased by its retention in the endoplamic reticulum (see accompanying paper), where a high concentration of chaperonins may enhance proper folding. It is unlikely, however, that the folding defect of m-GFP would manifest itself in the same way in systems where lower expression levels mean that aggregation may not occur to the same extent as in an *E. coli* overexpression system. Rather, it is more likely that partially or mis-folded apoprotein would, if it did not aggregate heavily, become a target of the cellular proteolytic machinery. Therefore, rather than becoming kinetically trapped by aggregation, improperly folded apoprotein would be depleted by degradation. In

support of this notion, accumulation in yeast of a GFP-nucleoplasmin fusion protein expressed from two gene copies steadily decreases with increasing incubation temperature.

Moreover, if the sensitivity of m-GFP apoprotein folding to temperature is a ubiquitous phenomenon, then use of the thermotolerant mutants described here should result in improved expression in a wide range of experimental systems. Indeed, in this work the inventors have demonstrated that the substitutions present in m-GFPA are capable of suppressing the thermosensitivity of m-GFP expression in the yeast *Saccharomyces cerevisiae*. Expression of m-GFPA has also been observed to give rise to significantly increased fluorescence in *Drosophila melanogaster* embryos incubated at 25°C (A Brand, personal communication) and expression of m-GFP5 fused to endoplasmic reticulum retention signals has been observed to result in high levels of fluorescence in *Arabidopsis thaliana* (data not shown). Most strikingly, expression of both m-GFPA and m-GFP5 has been found to result in greatly increased levels of fluorescence in mammalian cells. Therefore, we anticipate that the thermotolerant mutants described in this work and spectral variants derived from them will be of great benefit for expression in many experimental systems, particularly those such as mammalian cells that utilise higher incubation temperatures.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT:
(A) NAME: Medical Research Council
(B) STREET: 20 Park Crescent
(C) CITY: London
(E) COUNTRY: United Kingdom
(F) POSTAL CODE (ZIP): W1N 4AL
(G) TELEPHONE: (0171) 636 5422
(H) TELEFAX: (0171) 323 1331

(ii) TITLE OF INVENTION: Improvements in or Relating to Gene Expression

(iii) NUMBER OF SEQUENCES: 11

(iv) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGCGGATCCA AGGAGATATA ACAATGAGTA AAGGAGAAGA ACTTTTCACT 50

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GGCGAGCTCT TATTGTATA GTTCATCCAT GCC 33

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 122 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GATCATATGA AGCGGCACGA CTTCTTCAAG AGCGCCATGC CTGAGGGATA CGTGCAGGAG 60

AGGACCATCT TCTTCAAGGA CGACGGGAAC TACAAGACAC GTGCTGAAGT CAAGTTTGAG 120

GG 122

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 126 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GATGTATACG TTGTGGGAGT TGTAGTTGTA TTCCAACTTG TGGCCGAGGA TGTTTCCGTC 60

CTCCTTGAAA TCGATTCCCT TAAGCTCGAT CCTGTTGACG AGGGTGTCTC CCTCAAACTT 120

GACTTC 126

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GGCGGATCCA AGGAGATATA ACAATGAGTA AAGGAGAAGA ACTTTTCACT 50

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

GGCGAGCTCT TATTTGTATA GTTCATCCAT GCC 33

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 51 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GCCGAGCTCT TAGTGGTGGT GGTGGTGGTG TTTGTATAGT TCATCCATGC C 51

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 60 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGCGGATCCA AGGAGATATA ACAATGAAGA CTAATCTTTT TCTCTTTCTC ATCTTTTCAC 60

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 50 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GCCGAATTCG GCCGAGGATA ATGATAGGAG AAGTGAAAAG ATGAGAAAGA 50

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 792 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..789

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```
ATG AAG ACT AAT CTT TTT CTC TTT CTC ATC TTT TCA CTT CTC CTA TCA          48
Met Lys Thr Asn Leu Phe Leu Phe Leu Ile Phe Ser Leu Leu Leu Ser
 1               5                  10                  15

TTA TCC TCG GCC GAA TTC AGT AAA GGA GAA GAA CTT TTC ACT GGA GTT          96
Leu Ser Ser Ala Glu Phe Ser Lys Gly Glu Glu Leu Phe Thr Gly Val
            20                  25                  30

GTC CCA ATT CTT GTT GAA TTA GAT GGT GAT GTT AAT GGG CAC AAA TTT         144
Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe
        35                  40                  45

TCT GTC AGT GGA GAG GGT GAA GGT GAT GCA ACA TAC GGA AAA CTT ACC         192
Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr
    50                  55                  60

CTT AAA TTT ATT TGC ACT ACT GGA AAA CTA CCT GTT CCA TGG CCA ACA         240
Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr
65                  70                  75                  80

CTT GTC ACT ACT TTC TCT TAT GGT GTT CAA TGC TTT TCA AGA TAC CCA         288
Leu Val Thr Thr Phe Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro
                85                  90                  95

GAT CAT ATG AAG CGG CAC GAC TTC TTC AAG AGC GCC ATG CCT GAG GGA         336
Asp His Met Lys Arg His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly
            100                 105                 110

TAC GTG CAG GAG AGG ACC ATC TTC TTC AAG GAC GAC GGG AAC TAC AAG         384
Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys
        115                 120                 125

ACA CGT GCT GAA GTC AAG TTT GAG GGA GAC ACC CTC GTC AAC AGG ATC         432
Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile
    130                 135                 140

GAG CTT AAG GGA ATC GAT TTC AAG GAG GAC GGA AAC ATC CTC GGC CAC         480
Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His
145                 150                 155                 160

AAG TTG GAA TAC AAC TAC AAC TCC CAC AAC GTA TAC ATC ATG GCC GAC         528
Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp
                165                 170                 175

AAG CAG AAG AAC GGC ATC AAA GCC AAC TTC AAG ACC CGC CAC AAC ATC         576
Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Thr Arg His Asn Ile
            180                 185                 190

GAA GAC GGC GGC GTG CAA CTC GCT GAC CAT TAT CAA CAA AAT ACT CCA         624
Glu Asp Gly Gly Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro
        195                 200                 205
```

```
ATT GGC GAT GGC CCT GTC CTT TTA CCA GAC AAC CAT TAC CTG TCC ACA      672
Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr
    210                 215                 220

CAA TCT GCC CTT TCG AAA GAT CCC AAC GAA AAG AGA GAC CAC ATG GTC      720
Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val
225                 230                 235                 240

CTT CTT GAG TTT GTA ACA GCT GCT GGG ATT ACA CAT GGC ATG GAT GAA      768
Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu
                245                 250                 255

CTA TAC AAA CAC GAC GAA CTC TAA                                      792
Leu Tyr Lys His Asp Glu Leu
            260
```

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 263 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

```
Met Lys Thr Asn Leu Phe Leu Phe Leu Ile Phe Ser Leu Leu Leu Ser
 1               5                  10                  15

Leu Ser Ser Ala Glu Phe Ser Lys Gly Glu Glu Leu Phe Thr Gly Val
            20                  25                  30

Val Pro Ile Leu Val Glu Leu Asp Gly Asp Val Asn Gly His Lys Phe
        35                  40                  45

Ser Val Ser Gly Glu Gly Glu Gly Asp Ala Thr Tyr Gly Lys Leu Thr
    50                  55                  60

Leu Lys Phe Ile Cys Thr Thr Gly Lys Leu Pro Val Pro Trp Pro Thr
65                  70                  75                  80

Leu Val Thr Thr Phe Ser Tyr Gly Val Gln Cys Phe Ser Arg Tyr Pro
                85                  90                  95

Asp His Met Lys Arg His Asp Phe Phe Lys Ser Ala Met Pro Glu Gly
            100                 105                 110

Tyr Val Gln Glu Arg Thr Ile Phe Phe Lys Asp Asp Gly Asn Tyr Lys
        115                 120                 125

Thr Arg Ala Glu Val Lys Phe Glu Gly Asp Thr Leu Val Asn Arg Ile
    130                 135                 140

Glu Leu Lys Gly Ile Asp Phe Lys Glu Asp Gly Asn Ile Leu Gly His
145                 150                 155                 160
```

Lys Leu Glu Tyr Asn Tyr Asn Ser His Asn Val Tyr Ile Met Ala Asp
165 170 175

Lys Gln Lys Asn Gly Ile Lys Ala Asn Phe Lys Thr Arg His Asn Ile
180 185 190

Glu Asp Gly Gly Val Gln Leu Ala Asp His Tyr Gln Gln Asn Thr Pro
195 200 205

Ile Gly Asp Gly Pro Val Leu Leu Pro Asp Asn His Tyr Leu Ser Thr
210 215 220

Gln Ser Ala Leu Ser Lys Asp Pro Asn Glu Lys Arg Asp His Met Val
225 230 235 240

Leu Leu Glu Phe Val Thr Ala Ala Gly Ile Thr His Gly Met Asp Glu
245 250 255

Leu Tyr Lys His Asp Glu Leu
260

Claims

1. A DNA sequence encoding Green Fluorescent Protein (GFP), the sequence being modified relative to the wild type sequence so as to allow for more efficient expression in a plant cell of a functional GFP polypeptide.

2. A DNA sequence according to claim 1, wherein the modification is such as to reduce the probability of an RNA sequence transcribed therefrom being subject to erroneous splicing in a plant cell.

3. A DNA sequence according to claim 2, comprising a plurality of nucleotide substitutions relative to the wild type sequence, the substitutions serving to reduce the excision from the transcribed RNA of the portion corresponding to nucleotides 400-483 of the DNA sequence.

4. A DNA sequence according to claim 2 or 3, wherein the modification serves to decrease the A/U content of the transcribed RNA.

5. A DNA sequence according to any one of claims 2, 3 or 4, wherein the modification serves to decrease the A/U content of the portion of the transcribed RNA corresponding to nucleotides 400-483 of the DNA sequence.

6. A DNA sequence according to any one of the preceding claims, modified so as to cause an amino acid substitution at residue 163 and/or residue 175 relative to the wild type protein sequence.

7. A DNA sequence according to any one of claims 1-6, further comprising a cellular localisation signal directing the encoded GFP to a particular cellular compartment.

8. A DNA sequence according to claim 7, wherein the encoded GFP is directed to the endoplasmic reticulum (ER).

9. A DNA sequence according to claim 8, comprising the *Arabidopsis thaliana* basic chitinase localisation signal.

10. An RNA sequence capable of being transcribed from a DNA sequence according to any one of claims 1-9.

11. A modified GFP polypeptide comprising an amino acid substitution, relative to the wild type protein sequence, at residue 163 and/or 175.

12. A modified GFP according to claim 11, wherein residue 163 is alanine or a related amino acid.

13. A modified GFP according to claim 11 or 12, wherein residue 175 is glycine or a related amino acid.

14. A modified GFP according to any one of claims 11, 12 or 13, comprising one or more further amino acid substitutions relative to the wild type protein sequence.

15. A modified GFP according to any one of claims 11-14, comprising one or more further amino acid substitutions in, or immediately adjacent to, residues 65-67.

16. A modified GFP according to any one of claims 11-15, comprising a localisation signal.

17. A modified GFP according to claim 16, comprising a signal directing the GFP to the endoplasmic reticulum.

18. A modified GFP according to claim 17, comprising the *Arabidopsis thaliana* basic chitinase localisation signal.

19. A nucleic acid sequence encoding a polypeptide according to any one of claims 11-19.

20. A nucleic acid construct comprising a nucleic acid in accordacne with any one of claims 1-9 or claim 19.

21. An expression vector according to claim 20.

22. A construct according to claim 20 or 21, wherein the sequence encoding GFP is linked in frame to a sequence of interest.

23. A method of detecting the expression in a host cell of a sequence of interest, omprising causing the sequence of interest to be operably linked in frame with a ucleic acid sequence according to any one of claims 1-9 or claim 19, in a nucleic :id construct according to claim 20, introducing the construct into the host cell, and onitoring for fluorescence characterstic of the modified GFP.

A method of screening a plurality of host cells for those which have taken up a :leic acid sequence of interest, comprising mixing a construct according to claim 20 iprising the sequence of interest with the host cells in suitable conditions. ntaining the host cells under appropriate conditions for a sufficient length of time s to allow expression of modified GFP, and detecting those cells which exhibit ified GFP-mediated fluorescence.

Fig 1A

```
BamH I                                    Sac I
  |                                         |
ggatcc aaggagatataaca atg...GFP coding sequence...taa gagctc
cctagg ttcctctattattgt ..........................  ctcgag
         RBS
```

Fig. 1B

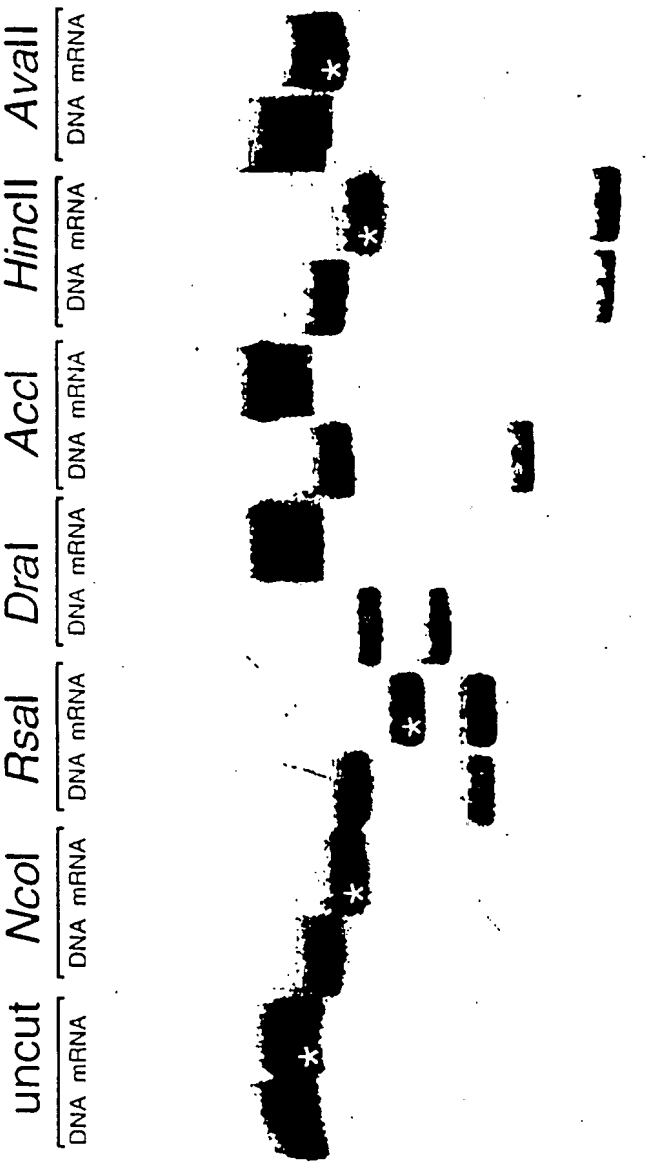

uncut
DNA mRNA
NcoI
DNA mRNA
RsaI
DNA mRNA
DraI
DNA mRNA
AccI
DNA mRNA
HincII
DNA mRNA
AvaII
DNA mRNA

Fig. 2A

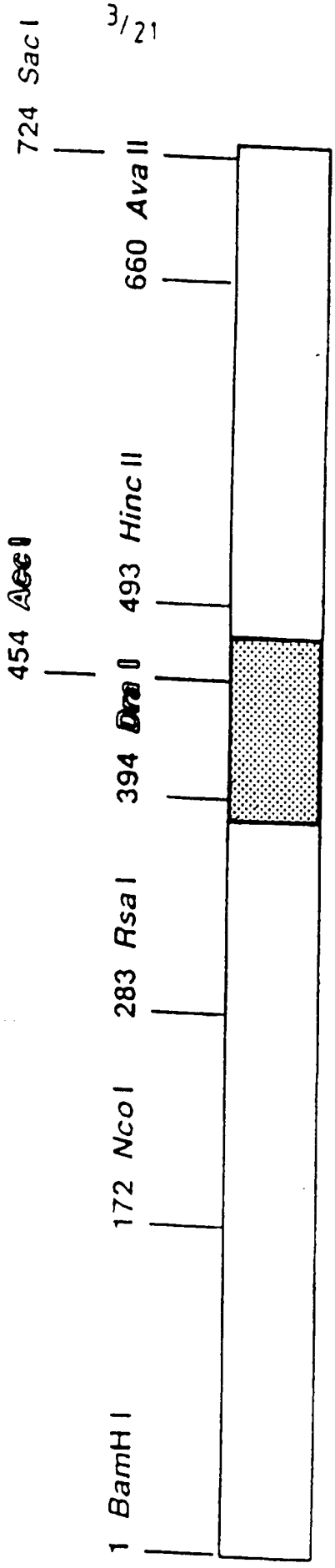

1 BamH I
172 Nco I
283 Rsa I
394 Dra I
454 Acc I
493 Hinc II
660 Ava II
724 Sac I
GFP coding sequence

Fig. 2B

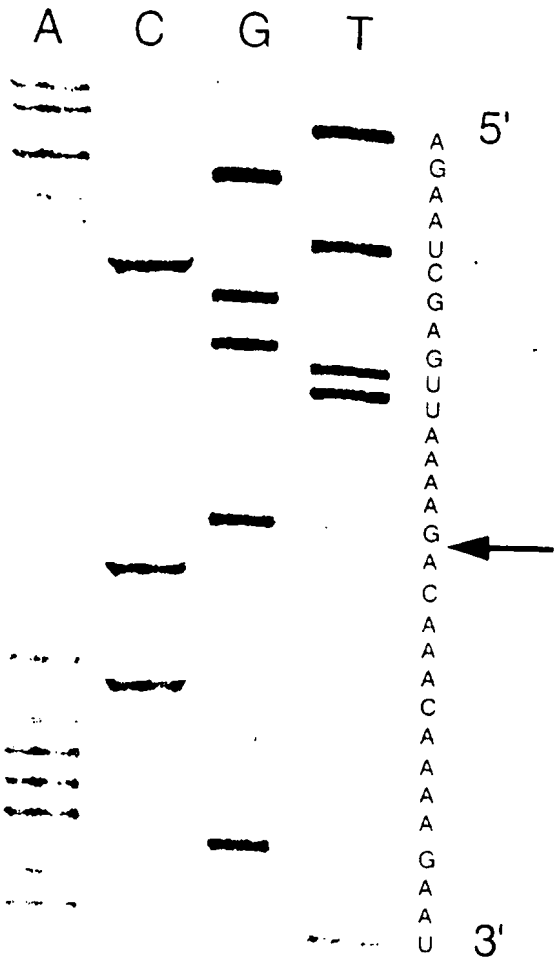

A
C
G
T
5'
AGAATUCGAGUUAAAAAGACAAAACAAAAAGAAU
3'

Fig 3A

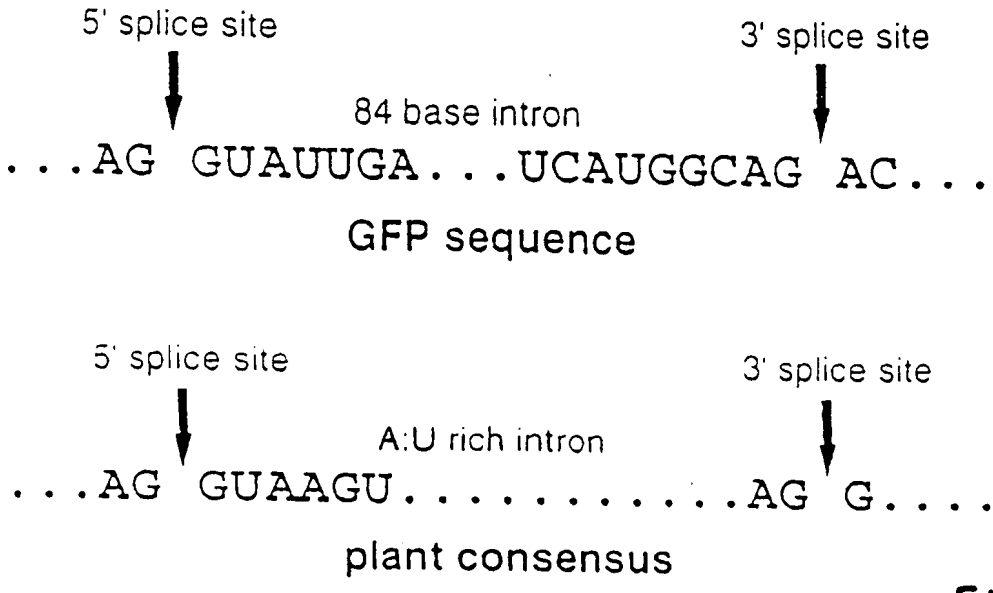

5' splice site
3' splice site
84 base intron
...AG GUAUUGA...UCAUGGCAG AC...
GFP sequence
5' splice site
3' splice site
A:U rich intron
...AG GUAAGU.....AG G....
plant consensus

Fig. 3B

201/61 231/71 *NdeI*

gtc act act ttc tct tat ggt gtt caa tgc ttt tca aga tac cca gat *cat atg* aaa cgg
gtc act act ttc tct tat ggt gtt caa tgc ttt tca aga tac cca gat *cat atg* aag cgg
V T T F S Y G V Q C F S R Y P D H M K R

261/81 291/91

cat gac ttt ttc aag agt gcc atg ccc gaa ggt tat gta cag gaa aga act ata ttt ttc
cac gac ttc ttc aag agc gcc atg cct gag gga tac gtg cag gag agg acc atc ttc ttc
H D F F K S A M P E G Y V Q E R T I F F

321/101 351/111

aaa gat gac ggg aac tac aag aca cgt gct gaa gtc aag ttt gaa *ggt gat acc ctt gtt*
aag gac gac ggg aac tac aag aca cgt gct gaa gtc aag ttt gag gga gac acc ctc gtc
K D D G N Y K T R A E V K F E G D T L V

381/121 411/131

aat aga atc gag tta aaa ggt att gat ttt aaa gaa gat gga aac att ctt gga cac aaa
aac agg atc gag ctt aag gga atc gat ttc aag gag gac gga aac atc ctc ggc cac aag
N R I E L K G I D F K E D G N I L G H K

441/141 *AccI* 471/151

ttg gaa tac aac tat aac tca cac aat gta tac atc atg gca gac aaa caa aag aat gga
ttg gaa tac aac tac aac tcc cac aac *gta tac* atc atg gca gac aaa caa aag aat gga
L E Y N Y N S H N V Y I M A D K Q K N G

top sequence = Aequoria victoria GFP, lower sequence = mGFP4
Intron sequences are underlined and the cryptic splice junctions are arrowed.
Mutated nucleotides are shown outlined. Nucleotide and amino acid
numbering starts at the initiation codon.

Fig. 4

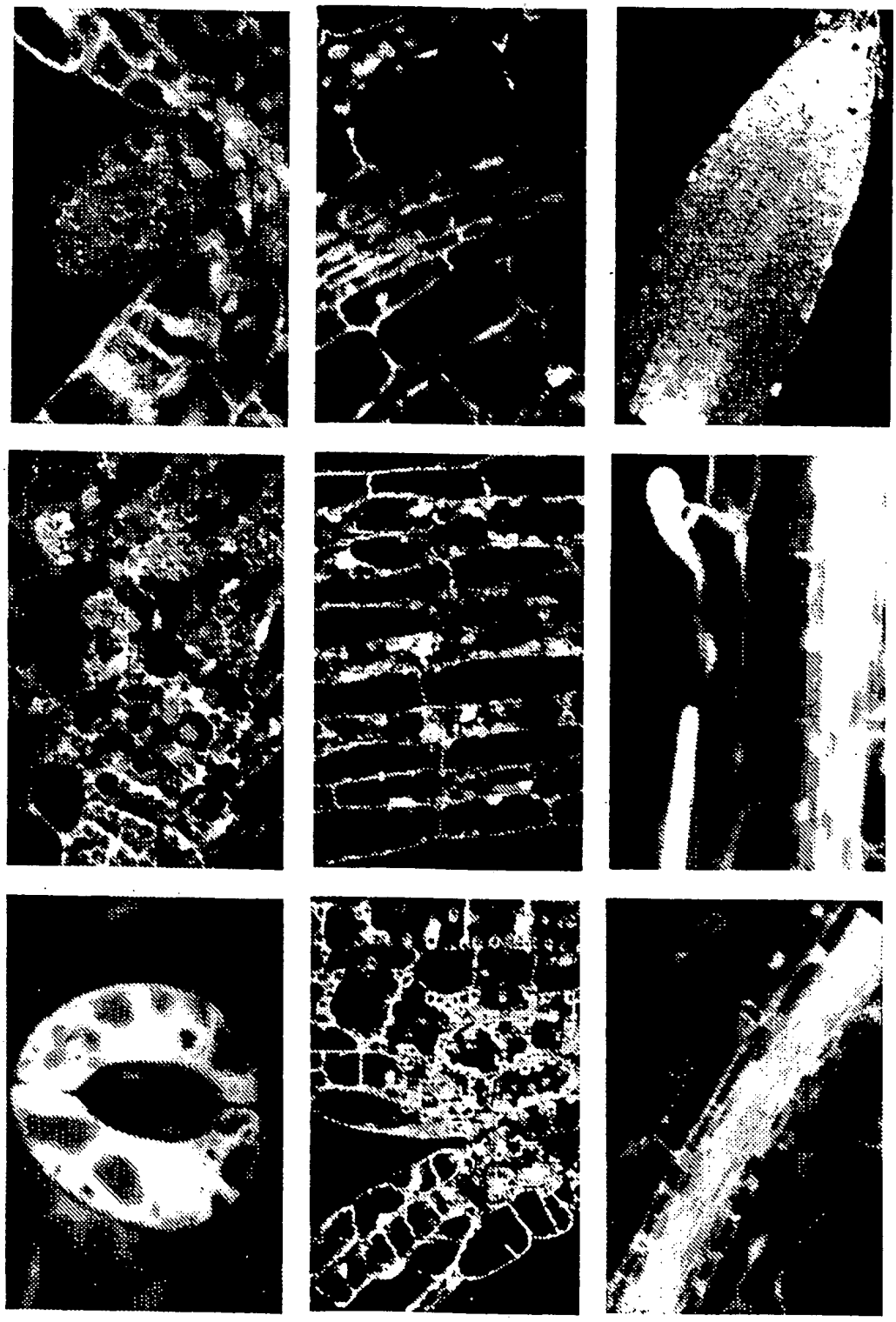
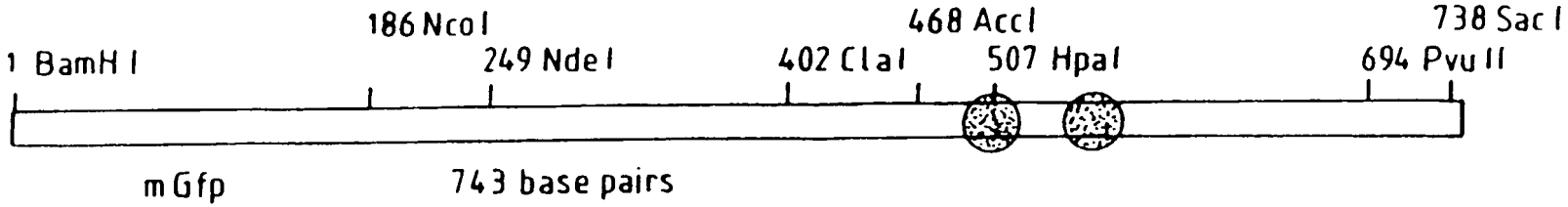


Fig. 5



501/161 **mGFP**

atc aaa gtt aac ttc aaa att aga cac aac att gaa gat gga agc gtt
ile lys val asn phe lys ile arg his asn ile glu asp gly ser val

501/161 **mGFP-B**

atc aaa gct aac ttc aaa att aga cac aac att gaa gat gga agc gtt
ile lys ala asn phe lys ile arg his asn ile glu asp gly ser val

501/161 **mGFP-A**

atc aaa gct aac ttc aaa att aga cac aac att gaa gat gga ggc gtt
ile lys ala asn phe lys ile arg his asn ile glu asp gly gly val

Fig. 6

WO 96/27675
PCT/GB96/00481
SUBSTITUTE SHEET (RULE 26)

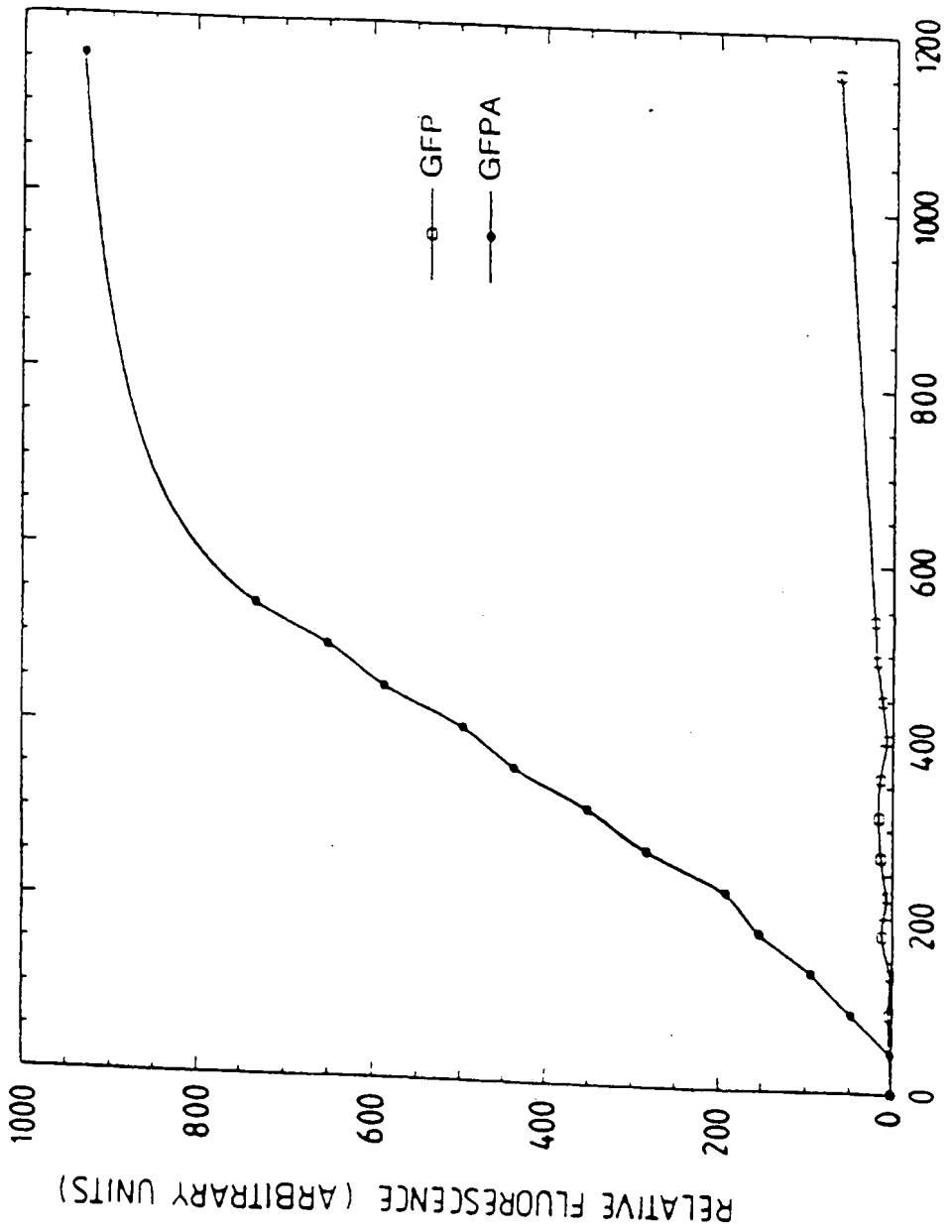

GFP
GFPA
RELATIVE FLUORESCENCE (ARBITRARY UNITS)
TIME (MIN)
0
200
400
600
800
1000
1200

Fig. 7

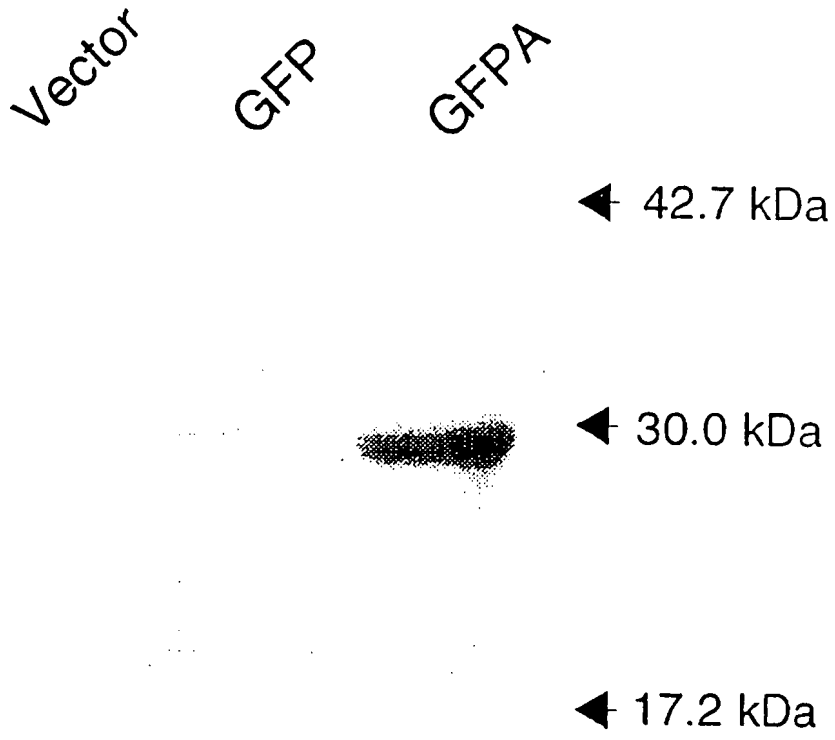


Fig. 8

Fig. 9

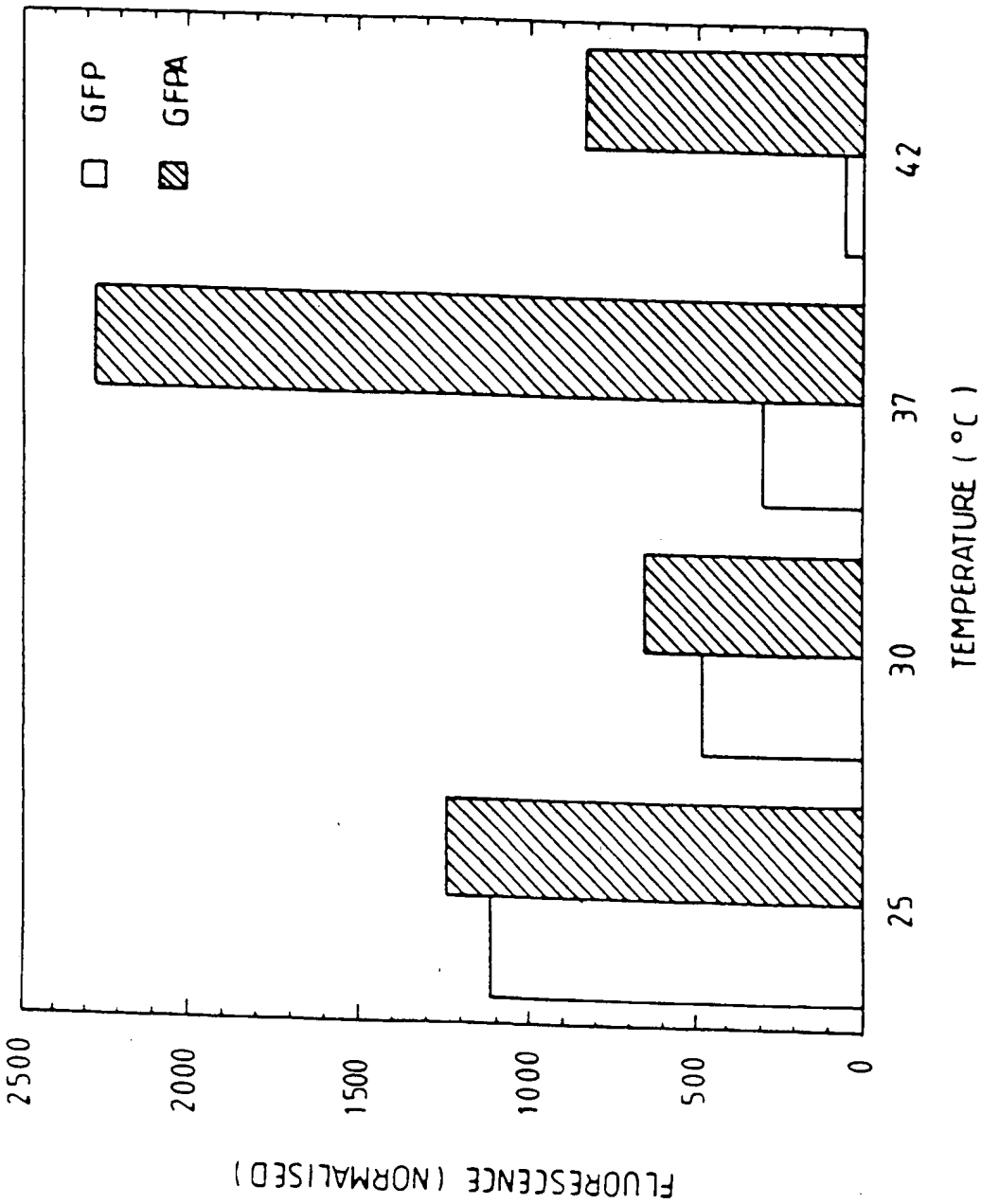
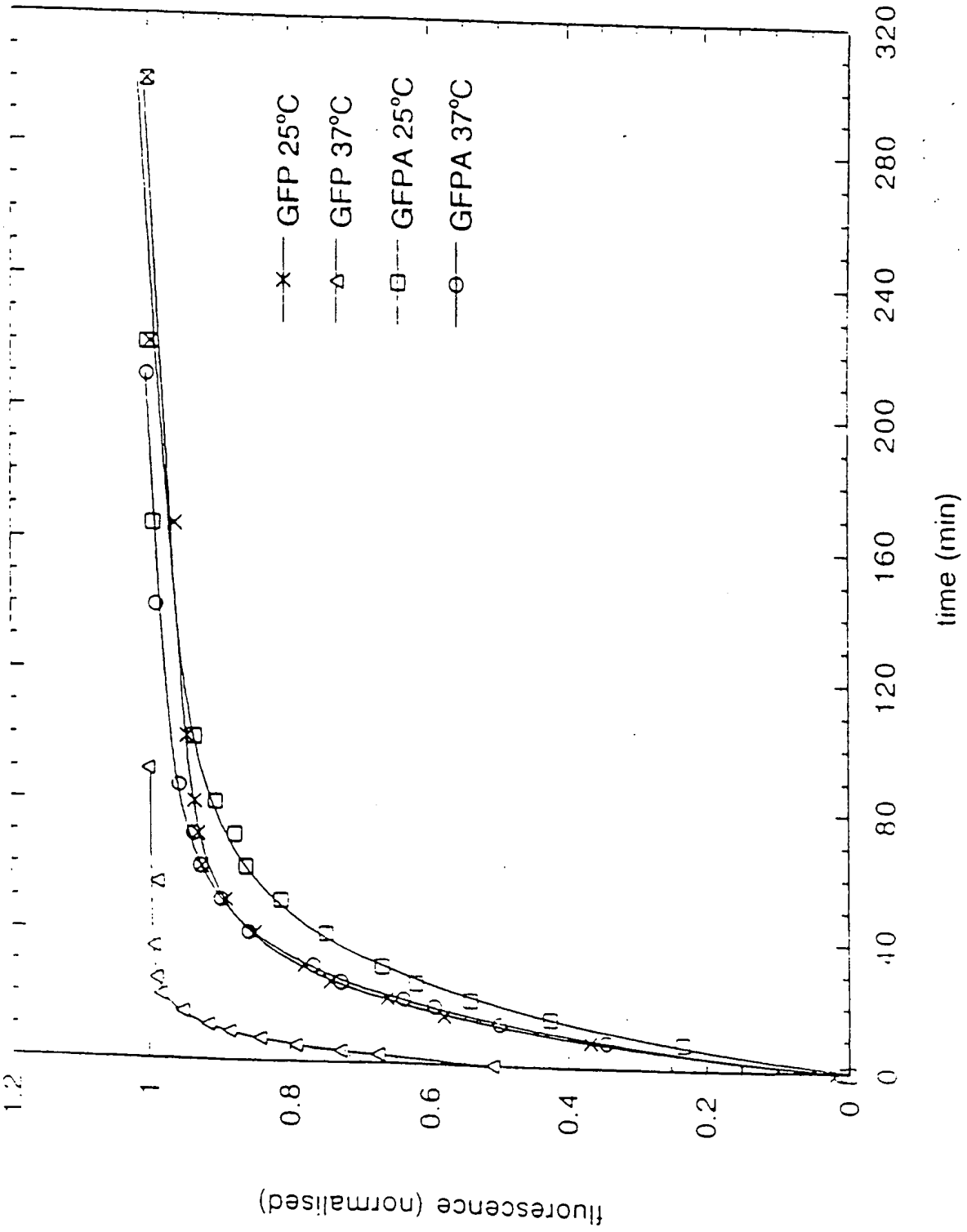

GFP
GFPA
2500
2000
1500
1000
500
0
FLUORESCENCE (NORMALISED)
25
30
37
42
TEMPERATURE (°C)

Fig.10


GFP 25°C
GFP 37°C
GFPA 25°C
GFPA 37°C
fluorescence (normalised)
time (min)
0
0.2
0.4
0.6
0.8
1
1.2
40
80
120
160
200
240
280
320

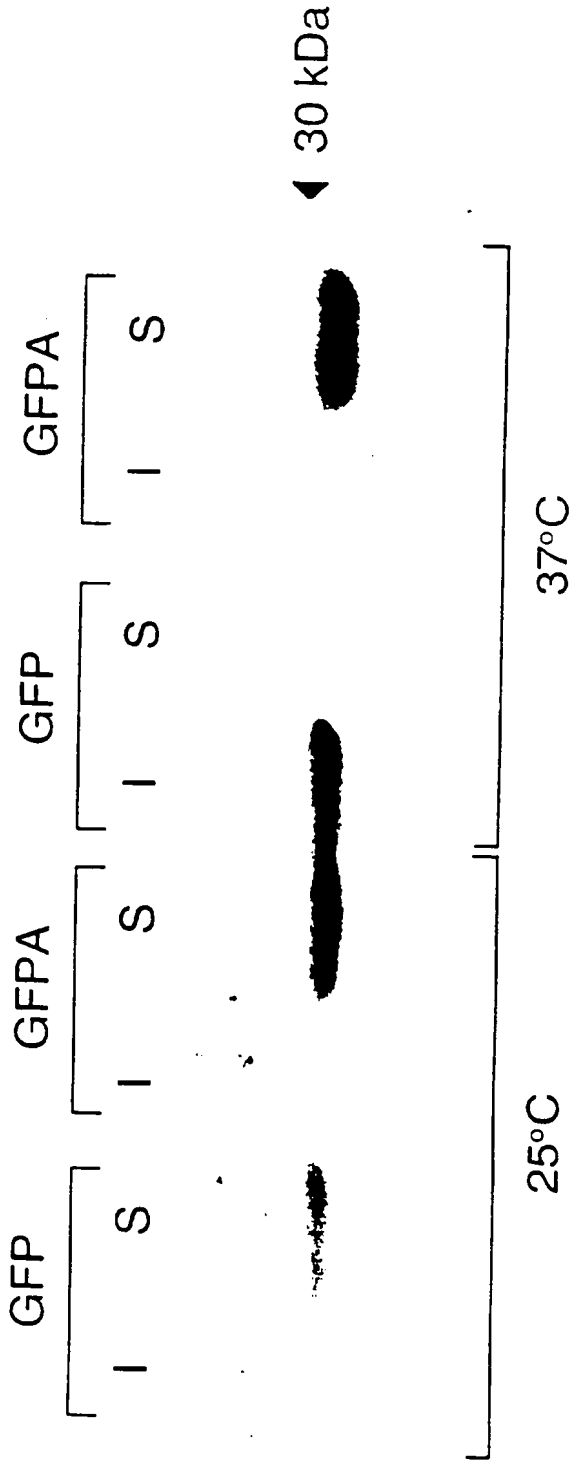
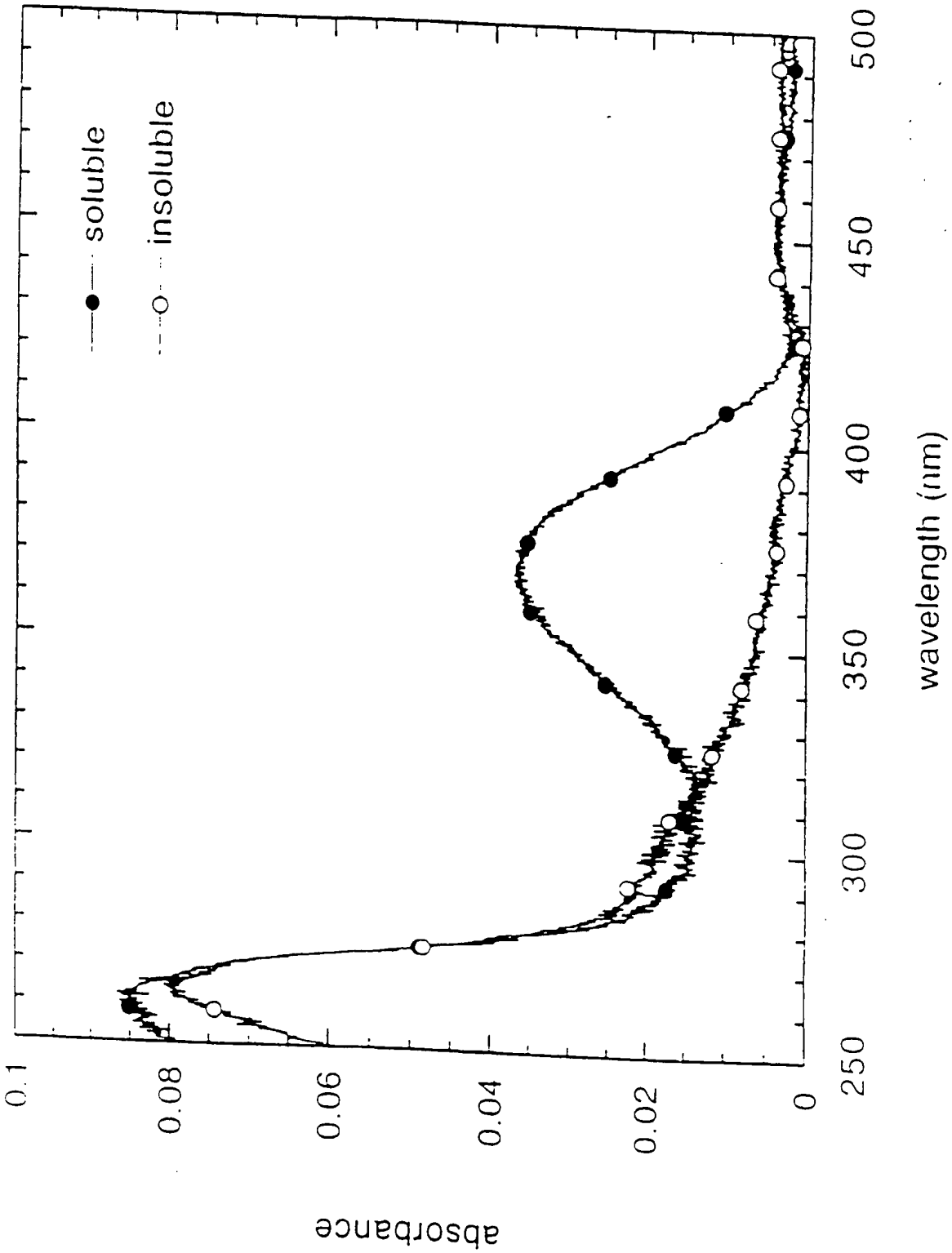


Fig. 11

Fig. 12


soluble
insoluble
absorbance
0.1
0.08
0.06
0.04
0.02
0
250
300
350
400
450
500
wavelength (nm)

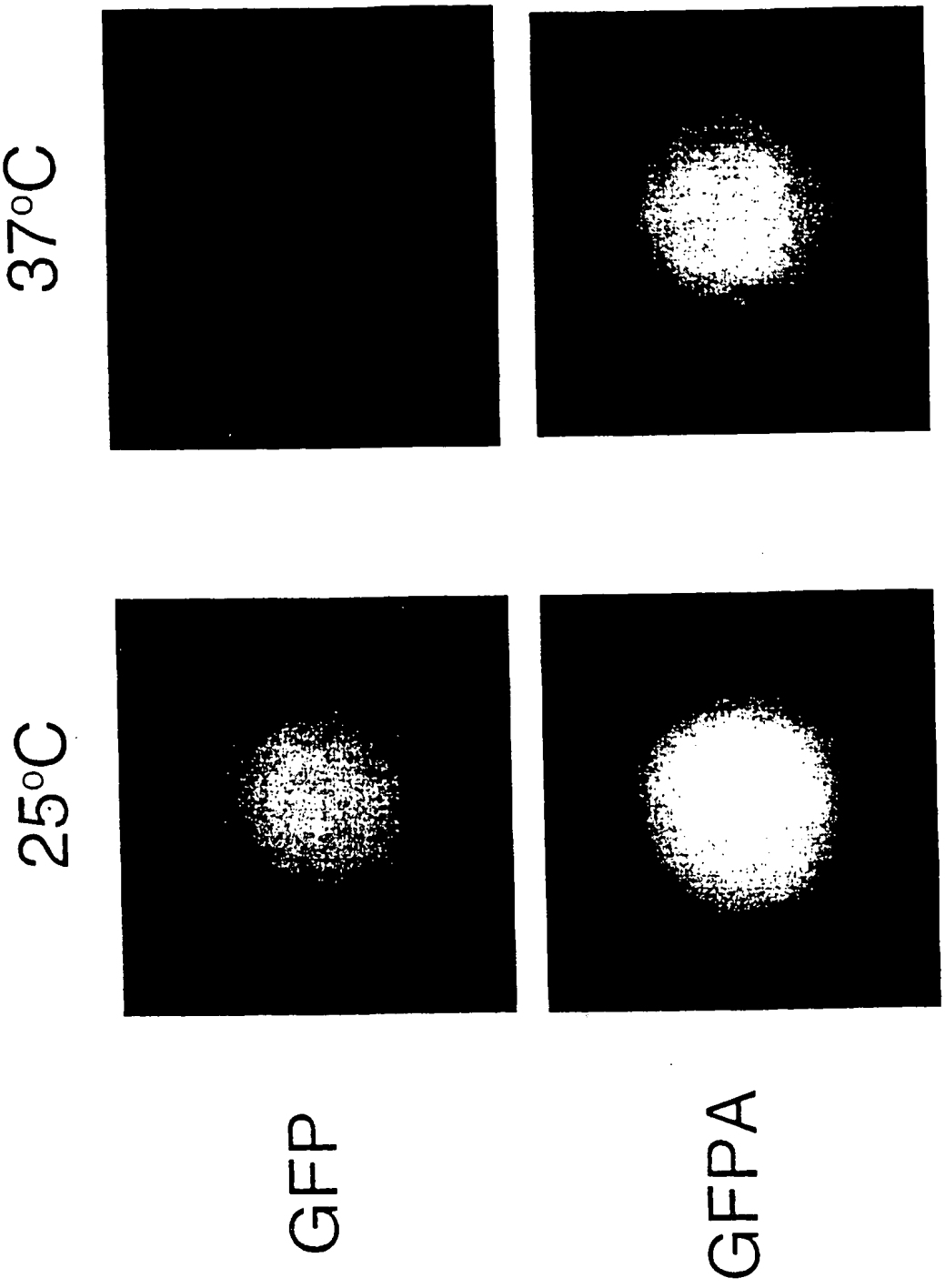
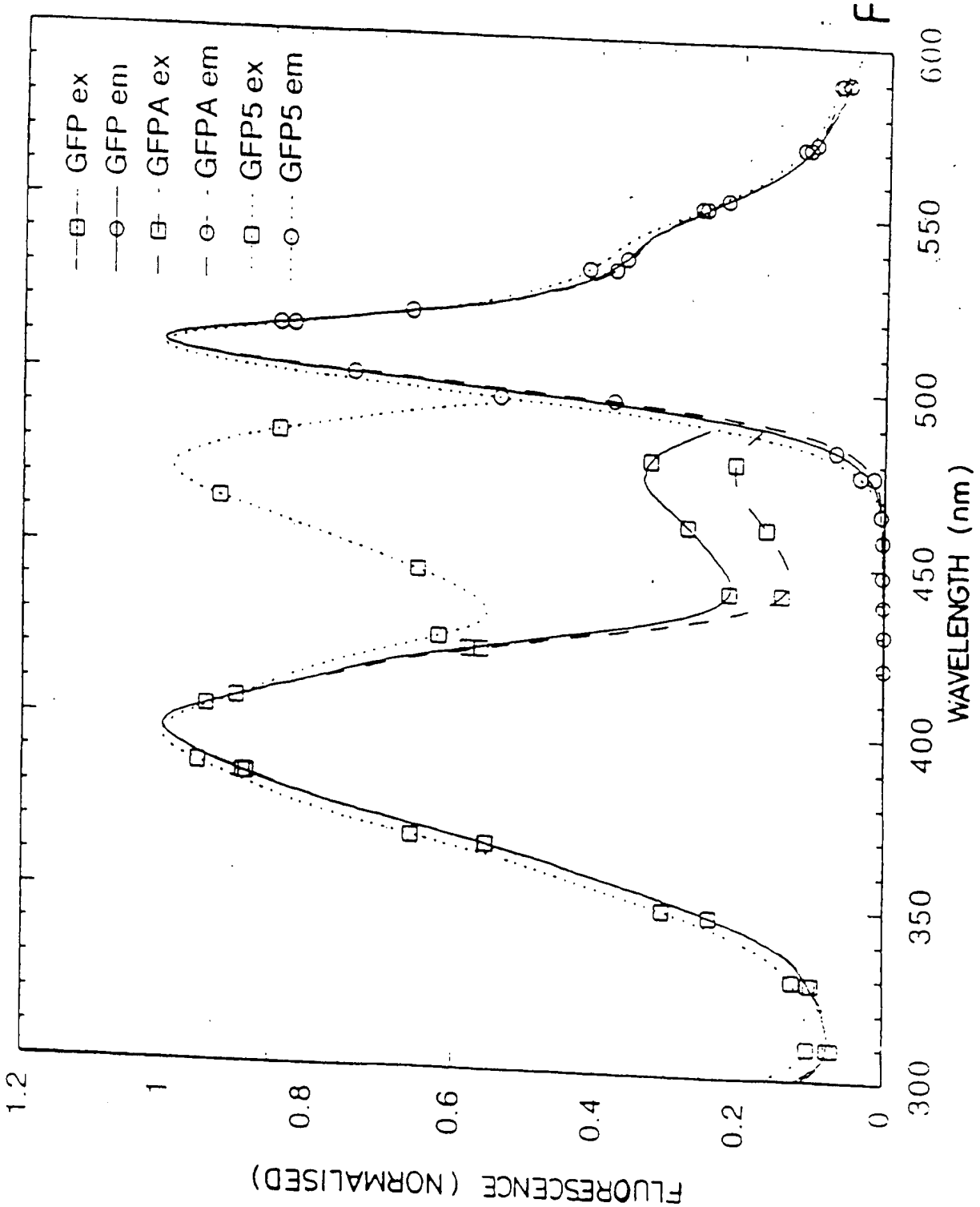


Fig. 13

Fig. 14


GFP ex
GFP em
GFPA ex
GFPA em
GFP5 ex
GFP5 em
FLUORESCENCE (NORMALISED)
1.2
1
0.8
0.6
0.4
0.2
0
300
350
400
450
500
550
600
WAVELENGTH (nm)

```
atg agt aaa gga gaa gaa ctt ttc act gga gtt gtc cca att ctt gtt gaa tta gat    gfp
atg agt aaa gga gaa gaa ctt ttc act gga gtt gtc cca att ctt gtt gaa tta gat    m-gfp5
 M   S   K   G   E   E   L   F   T   G   V   V   P   I   L   V   E   L   D

ggt gat gtt aat ggg cac aaa ttt tct gtc agt gga gag ggt gaa ggt gat gca aca tac    gfp
ggt gat gtt aat ggg cac aaa ttt tct gtc agt gga gag ggt gaa ggt gat gca aca tac    m-gfp5
 G   D   V   N   G   H   K   F   S   V   S   G   E   G   E   G   D   A   T   Y

                                                        Nco I
gga aaa ctt acc ctt aaa ttt att tgc act act gga aaa cta cct gtt cca tgg cca aca    gfp
gga aaa ctt acc ctt aaa ttt att tgc act act gga aaa cta cct gtt cca tgg cca aca    m-gfp5
 G   K   L   T   L   K   F   I   C   T   T   G   K   L   P   V   P   W   P   T

                                                            Nde I
ctt gtc act act ttc tct tat ggt gtt caa tgc ttt tca aga tac cca gat cat atg aaa    gfp
ctt gtc act act ttc tct tat ggt gtt caa tgc ttt tca aga tac cca gat cat atg aag    m-gfp5
 L   V   T   T   F   S   Y   G   V   Q   C   F   S   R   Y   P   D   H   M   K

cgg cat gac ttt ttc aag agt gcc atg ccc gaa ggt tat gta cag gaa aga act ata ttt    gfp
cgg cac gac ttc ttc aag agc gcc atg cct gag gga tac gtg cag gag agg acc atc ttc    m-gfp5
 R   H   D   F   F   K   S   A   M   P   E   G   Y   V   Q   E   R   T   I   F

ttc aaa gat gac ggg aac tac aag aca cgt gct gaa gtc aag ttt gaa ggt gat acc ctt    gfp
ttc aag gac gac ggg aac tac aag aca cgt gct gaa gtc aag ttt gag gga gac acc ctc    m-gfp5
 F   K   D   D   G   N   Y   K   T   R   A   E   V   K   F   E   G   D   T   L

                                                cryptic intron
gtt aat aga atc gag tta aaa ggt att gat ttt aaa gaa gat gga aac att ctt gga cac    gfp
gtc aac agg atc gag ctt aag gga atc gat ttc aag gag gac gga aac atc ctc ggc cac    m-gfp5
 V   N   R   I   E   L   K   G   I   D   F   K   E   D   G   N   I   L   G   H
```

Fig.15 Sheet 1

Acc I

aaa ttg gaa tac aac tat aac tca cac aat gta tac atc atg gca gac aaa caa aag aat *gfp*
aag ttg gaa tac aac tac aac tcc cac aac gta tac atc atg gcc gac aag cag aag aac *m-gfp5*
K L E Y N Y N S H N V Y I M A D K Q K N

V I S

gga atc aaa gtt aac ttc aaa att aga cac aac att gaa gat gga agc gtt caa cta gca *gfp*
ggc atc aaa gcc aac ttc aag acc cgc cac aac atc gaa gac ggc ggc gtg caa ctc gct *m-gfp5*
G I K A N F K T R H N I E D G G V Q L A

gac cat tat caa caa aat act cca att ggc gat ggc cct gtc ctt tta cca gac aac cat *gfp*
gat cat tat caa caa aat act cca att ggc gat ggc cct gtc ctt tta cca gac aac cat *m-gfp5*
D H Y Q Q N T P I G D G P V L L P D N H

tac ctg tcc aca caa tct gcc ctt tcg aaa gat ccc aac gaa aag aga gac cac atg gtc *gfp*
tac ctg tcc aca caa tct gcc ctt tcg aaa gat ccc aac gaa aag aga gac cac atg gtc *m-gfp5*
Y L S T Q S A L S K D P N E K R D H M V

ctt ctt gag ttt gta aca gct gct ggg att aca cat ggc atg gat gaa cta tac aaa taa *gfp*
ctt ctt gag ttt gta aca gct gct ggg att aca cat ggc atg gat gaa cta tac aaa taa *m-gfp5*
L L E F V T A A G I T H G M D E L Y K •

Fig. 15 Sheet 2

```
21/1                                                    51/11
ATG AAG ACT AAT CTT TTT CTC TTT CTC ATC TTT TCA CTT CTC CTA TCA TTA TCC TCG GCC
M   K   T   N   L   F   L   F   L   I   F   S   L   L   L   S   L   S   S   A
81/21                                                   111/31
GAA TTC agt aaa gga gaa gaa ctt ttc act gga gtt gtc cca att ctt gtt gaa tta gat
E   F   S   K   G   E   E   L   F   T   G   V   V   P   I   L   V   E   L   D
141/41                                                  171/51
ggt gat gtt aat ggg cac aaa ttt tct gtc agt gga gag ggt gaa ggt gat gca aca tac
G   D   V   N   G   H   K   F   S   V   S   G   E   G   E   G   D   A   T   Y
201/61                                                  231/71
gga aaa ctt acc ctt aaa ttt att tgc act act gga aaa cta cct gtt cca tgg cca aca
G   K   L   T   L   K   F   I   C   T   T   G   K   L   P   V   P   W   P   T
261/81                                                  291/91
ctt gtc act act ttc tct tat ggt gtt caa tgc ttt tca aga tac cca gat cat atg aag
L   V   T   T   F   S   Y   G   V   Q   C   F   S   R   Y   P   D   H   M   K
321/101                                                 351/111
cgg cac gac ttc ttc aag agc gcc atg cct gag gga tac gtg cag gag agg acc atc ttc
R   H   D   F   F   K   S   A   M   P   E   G   Y   V   Q   E   R   T   I   F
381/121                                                 411/131
ttc aag gac gac ggg aac tac aag aca cgt gct gaa gtc aag ttt gag gga gac acc ctc
F   K   D   D   G   N   Y   K   T   R   A   E   V   K   F   E   G   D   T   L
```

Fig. 16 Sheet 1

```
441/141                                                 471/151
gtc aac agg atc gag ctt aag gga atc gat ttc aag gag gac gga aac atc ctc gga cac
V   N   R   I   E   L   K   G   I   D   F   K   E   D   G   N   I   L   G   H
501/161                                                 531/171
aag ttg gaa tac aac tac aac tcc cac aac gta tac atc atg gcc gac aag cag aag aac
K   L   E   Y   N   Y   N   S   H   N   V   Y   I   M   A   D   K   Q   K   N
561/181                                                 591/191
ggc atc aaa gcc aac ttc aag acc cgc cac aac atc gaa gac ggc ggc gtg caa ctc gct
G   I   K   A   N   F   K   T   R   H   N   I   E   D   G   G   V   Q   L   A
621/201                                                 651/211
gat cat tat caa caa aat act cca att ggc gat ggc cct gtc ctt tta cca gac aac cat
D   H   Y   Q   Q   N   T   P   I   G   D   G   P   V   L   L   P   D   N   H
681/221                                                 711/231
tac ctg tcc aca caa tct gcc ctt tcg aaa gat ccc aac gaa aag aga gac cac atg gtc
Y   L   S   T   Q   S   A   L   S   K   D   P   N   E   K   R   D   H   M   V
741/241                                                 771/251
ctt ctt gag ttt gta aca gct gct ggg att aca cat ggc atg gat gaa cta tac aaa cac
L   L   E   F   V   T   A   A   G   I   T   H   G   M   D   E   L   Y   K   H
801/261
gac gaa ctc taa
D   E   L   *
```

Fig. 16 Sheet 2

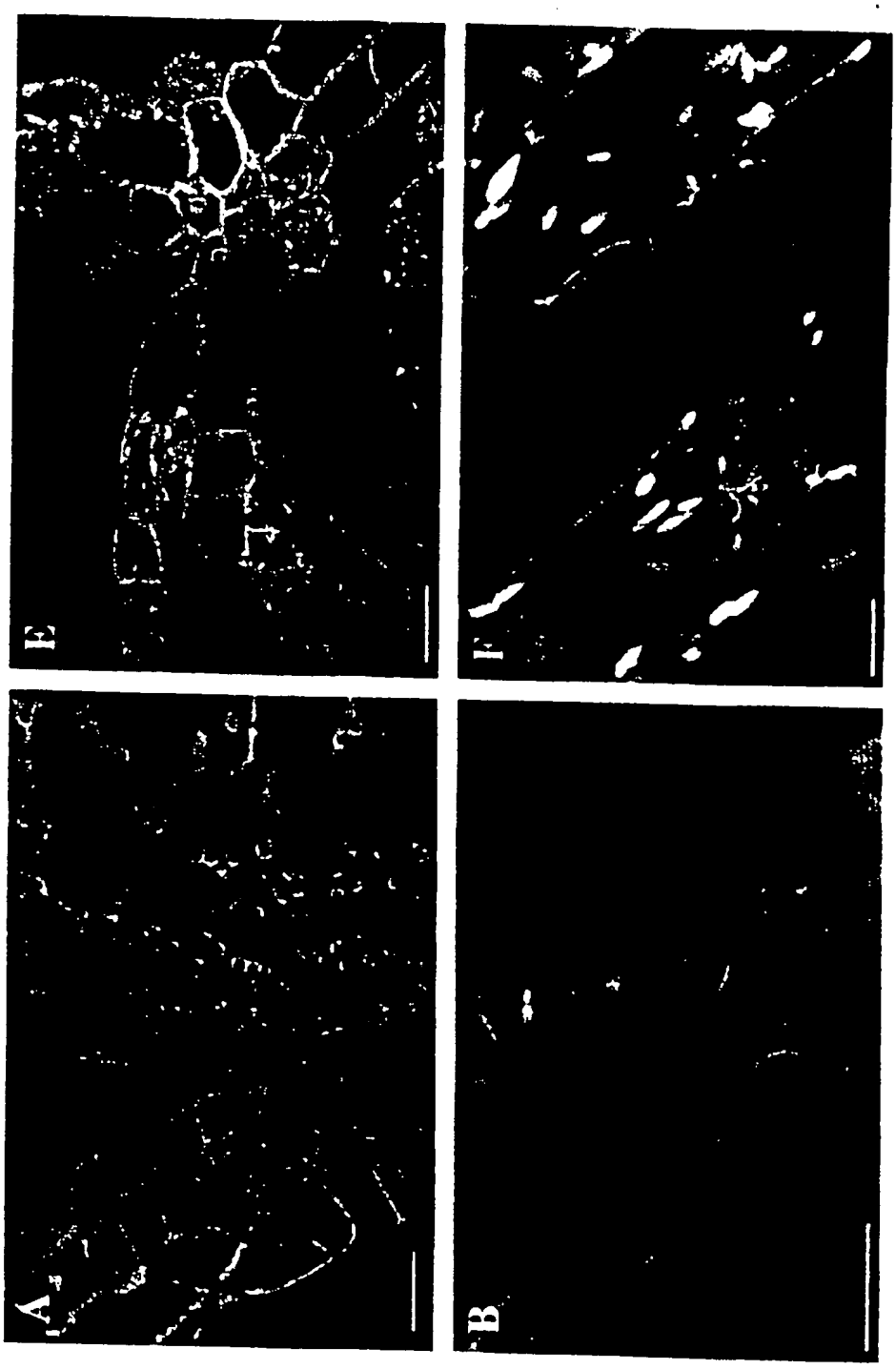


Fig. 17 Sheet 1

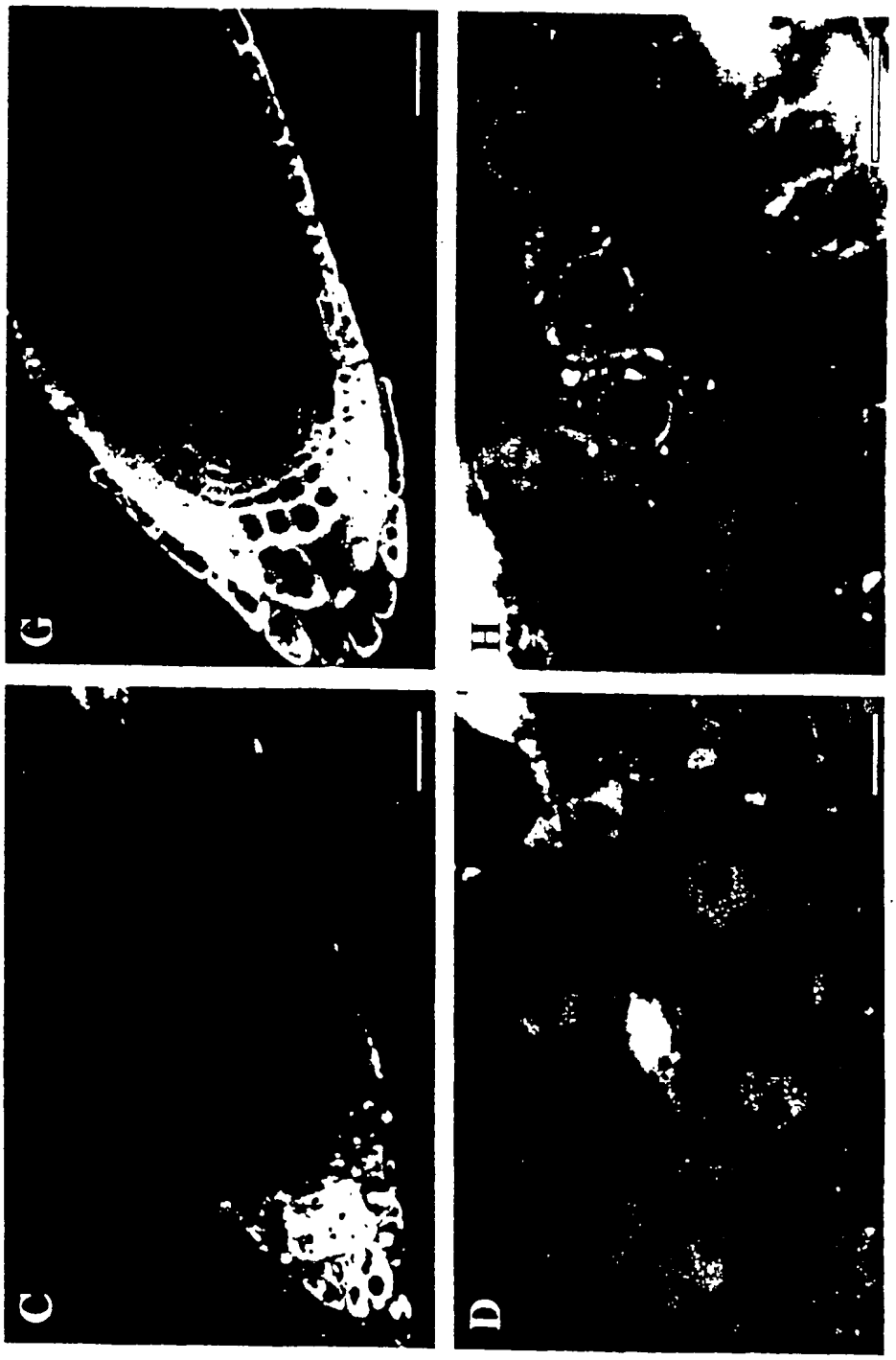


Fig. 17 Sheet 2

INTERNATIONAL SEARCH REPORT

Inte nal Application No
PCT/GB 96/00481

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/82 C12N15/12 C07K14/435 G01N33/52 G01N33/53

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K G01N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	TRENDS IN GENETICS, vol. 11, no. 8, August 1995, pages 328-329, XP002003595 HASELOFF, J., ET AL.: "GFP in plants" see the whole document ---	1-5,10, 20,21,24
P,X	CURR. BIOL. (1996), 6(3), 325, pages 325-330, XP000571865 CHUI, W., ET AL.: "Engineered GFP as a vital reporter plants" see the whole document ---	1-5,10, 20,21,24
P,X	WO,A,95 07463 (UNIV COLUMBIA ;WOODS HOLE OCEANOGRAPHIC INST (US); CHALFIE MARTIN) 16 March 1995 see the whole document ---	1
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

2

Date of the actual completion of the international search	Date of mailing of the international search report
28 May 1996	07.06.96
Name and mailing address of the ISA European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl, Fax (+ 31-70) 340-3016	Authorized officer Maddox, A

Form PCT ISA 210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P,X	PLANT CELL REPORTS, vol. 14, April 1995, pages 403-406, XP000571886 NIEDZ, R.P., ET AL.: "Green fluorescent protein : an in vivo reporter of plant gene expression" see the whole document	23,24
P,X	THE PLANT JOURNAL, vol. 8, no. 5, November 1995, pages 777-784, XP002003596 SHEEN, J., ET AL.: "Green-fluorescent protein as a new vital marker in plant cells" see the whole document	23,24
A	SCIENCE, vol. 263, 11 February 1994, pages 802-805, XP002003599 CHLAFIE, M., ET AL.: "Green fluorescent protein as a marker for gene expression" see the whole document	1-23
A	WO,A,91 01305 (UNIV WALES MEDICINE) 7 February 1991 see claim 10	1-23
A	NATURE, vol. 369, June 1994, pages 400-403, XP002003600 WANG, S., ET AL.: "Implications for bcd mRNA localization from spatial ditribution of exu protein in Drosophila oogenesis" see the whole document	23

INTERNATIONAL SEARCH REPORT

Inter nal Application No
PCT/GB 96/00481

Patent document cited in search report	Publication date	Patent family member(s)		Publication date
WO-A-9507463	16-03-95	US-A-	5491084	13-02-96
		AU-B-	7795794	27-03-95
		CA-A-	2169298	16-03-95
WO-A-9101305	07-02-91	AU-B-	6054590	22-02-91
		EP-A-	0484369	13-05-92
		JP-T-	5501862	08-04-93